

(19)日本国特許庁 ( J P )

(12) 公 開 特 許 公 報 ( A )

(11)特許出願公開番号

特開平10-27182

(43)公開日 平成10年(1998) 1 月27日

(51)Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30			G 0 6 F 15/403	3 4 0 Z
3/14	3 6 0		3/14	3 6 0 C
			15/40	3 1 0 F
				3 7 0 A

審査請求 未請求 請求項の数20 F D (全 42 頁)

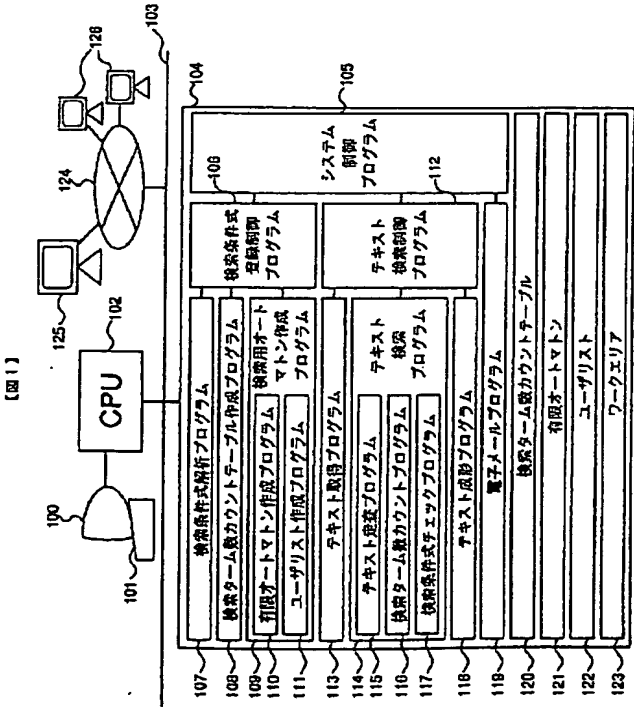
(21)出願番号	特願平8-201348	(71)出願人	000005108 株式会社日立製作所 東京都千代田区神田駿河台四丁目 6 番地
(22)出願日	平成 8 年(1996) 7 月11日	(72)発明者	菅谷 奈津子 神奈川県川崎市幸区鹿島田890番地 株式 会社日立製作所情報・通信開発本部内
		(72)発明者	川口 久光 神奈川県川崎市幸区鹿島田890番地 株式 会社日立製作所情報・通信開発本部内
		(72)発明者	山崎 紀之 神奈川県横浜市戸塚区戸塚町5030番地 株 式会社日立製作所ソフトウェア開発本部内
		(74)代理人	弁理士 笹岡 茂 (外 1 名)

(54)【発明の名称】 文書検索配送方法および装置

(57)【要約】

【課題】 ユーザ数が増えた場合でも、ユーザ数に依らない高速なテキスト検索配布機能を実現する文書検索配送システムを提供する。

【解決手段】 検索条件式を解析し、検索条件式に含まれる検索ターム（以下、ターム）数を検索ターム数カウントテーブル（以下、テーブルA）に格納し、検索条件式に含まれる全てのタームを照合する有限オートマトンを作成し、検索条件式に含まれるタームがどのユーザのものであるかを記憶したユーザ識別子テーブルを作成し、対象テキストを有限オートマトンで走査してタームを照合し、ユーザ識別子テーブルを参照してテキスト中にタームと一致する部分文字列が出現したターム数をユーザ毎に算出し、テーブルAの算出したターム数用格納領域に格納し、テーブルA内の算出したターム数と検索条件式中のターム数とを比較し、比較結果が一致している場合は、そのユーザに対し該テキストを配布する。



**【特許請求の範囲】**

【請求項 1】 一つ以上の情報源から入手した文書情報のテキストデータを対象として、

1 個以上の検索タームを含む 1 人以上のユーザが指定した検索条件式を登録する検索条件式登録ステップと、  
テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布ステップを有する文書検索配送方法において、  
前記テキスト検索配布ステップは、前記テキストをただ一度走査することによって前記複数の検索条件式の該テキストに対する成否を判断するテキスト検索ステップを有することを特徴とする文書検索配送方法。

【請求項 2】 請求項 1 記載の文書検索配送方法において、

前記検索条件式登録ステップは、  
前記検索条件式から全ての検索タームを抽出する検索条件式解析ステップと、  
ユーザ毎にユーザと該ユーザの検索条件式から抽出された全ての検索タームの数を含む管理情報を格納する検索ターム数カウントテーブルを作成する検索ターム数カウントテーブル作成ステップと、  
前記検索条件式から抽出した全ての検索タームを、テキストのただ一度の走査により照合する際に参照する多重文字列照合テーブルを生成する多重文字列照合テーブル生成ステップと、  
検索条件式から抽出された各検索ターム対応に該検索条件式を指定したユーザのユーザ識別子をリストとしてつないだユーザリストを生成するユーザリスト生成ステップを有し、

前記テキスト検索配布ステップは、  
該テキストに対する前記検索条件式の成否の判断時に、  
前記多重文字列照合テーブルを参照して該テキストを走査することによって、前記検索条件式解析ステップにより抽出された全ての検索タームを照合するテキスト走査ステップと、  
前記テキスト走査ステップによって照合された検索タームと前記ユーザリストと前記検索ターム数カウントテーブルを照合することにより、該テキストに対する前記検索条件式の成否を判断する検索条件式成否判断ステップを有することを特徴とする文書検索配送方法。

【請求項 3】 請求項 2 記載の文書検索配送方法において、

前記多重文字列照合テーブルとして有限オートマトンを用いることを特徴とする文書検索配送方法。

【請求項 4】 請求項 2 または請求項 3 記載の文書検索配送方法において、

前記検索条件式成否判断ステップは、  
前記ユーザリストを参照し、前記テキスト走査ステップによって照合された検索タームの個数をユーザ毎に算出

する検索ターム照合数算出ステップと、

前記検索ターム照合数算出ステップで算出された検索タームの個数と前記検索ターム数カウントテーブルに格納されている検索タームの個数とを比較し、一致している場合には該検索タームが含まれる検索条件式が成立しているものとみなす検索ターム数比較ステップを有することを特徴とする文書検索配送方法。

【請求項 5】 一つ以上の情報源から入手した文書情報のテキストデータを対象として、

1 個以上の検索タームを含む 1 人以上のユーザが指定した検索条件式を登録する検索条件式登録ステップと、  
テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布ステップを有する文書検索配送方法において、  
1 人以上のユーザあるいはシステム管理者が指定したテキスト配布の条件を記した配布条件を含む配布条件設定式を登録する配布条件設定式登録ステップを有し、  
前記テキスト検索配布ステップは、  
前記テキストをただ一度走査することによって前記複数の検索条件式の該テキストに対する成否を判断するテキスト検索ステップと、  
前記テキスト検索ステップによって前記検索条件式が成立したユーザに対して、前記配布条件設定式登録ステップによって登録された前記配布条件が成立した時点で前記テキストを配布するテキスト配布制御ステップを有することを特徴とする文書検索配送方法。

【請求項 6】 請求項 5 記載の文書検索配送方法において、

前記配布条件設定式登録ステップは、  
前記配布条件設定式から配布条件を設定すべきユーザの識別子と配布条件を抽出する配布条件設定式解析ステップと、  
前記配布条件設定式解析ステップにおいて前記配布条件設定式から抽出されたユーザの識別子と配布条件を格納した配布条件管理テーブルを作成する配布条件管理テーブル作成ステップを有し、

前記テキスト配布制御ステップは、  
前記配布条件管理テーブルを参照して前記配布条件の成否を判断する配布条件成否判断ステップと、  
前記配布条件成否判断ステップによって前記配布条件が成立していると判断された時点でユーザに対して前記テキストを配布するテキスト配布ステップを有することを特徴とする文書検索配送方法。

【請求項 7】 請求項 6 記載の文書検索配送方法において、

前記配布条件として、配布する時間、配布する件数またはテキスト検索から配布までの遅延時間を用いることを特徴とする文書検索配送方法。

【請求項 8】 一つ以上の情報源から入手した文書情報

のテキストデータを対象として、

1 個以上の検索タームを含む 1 人以上のユーザが指定した検索条件式を登録する検索条件式登録ステップと、  
テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布ステップを有する文書検索配送方法において、

前記検索条件式の削除が指示された場合には該検索条件式を削除する検索条件式削除ステップを有することを特徴とした文書検索配送方法。

【請求項 9】 請求項 8 記載の文書検索配送方法において、

前記検索条件式登録ステップは、

前記検索条件式から全ての検索タームを抽出する検索条件式解析ステップと、

ユーザ毎にユーザと該ユーザの検索条件式から抽出された全ての検索タームの数を含む管理情報を格納する検索ターム数カウントテーブルを作成する検索ターム数カウントテーブル作成ステップと、

前記検索条件式から抽出した全ての検索タームを、テキストのただ一度の走査により照合する際に参照する多重文字列照合テーブルを生成する多重文字列照合テーブル生成ステップと、

検索条件式から抽出された各検索ターム対応に該検索条件式を指定したユーザのユーザ識別子をリストとしてつないだユーザリストを生成するユーザリスト生成ステップを有し、

前記検索条件式削除ステップは、

削除が指示された前記検索条件式に関連する情報を前記検索ターム数カウントテーブルおよび前記ユーザリストから削除する検索条件式管理テーブル削除ステップを有することを特徴とする文書検索配送方法。

【請求項 10】 請求項 9 記載の文書検索配送方法において、

前記検索条件式登録ステップは、

さらに、前記検索条件式解析ステップにより抽出された検索タームを格納した検索ターム管理テーブルを作成する検索ターム管理テーブル作成ステップを有し、

前記検索条件式管理テーブル削除ステップは、

前記検索ターム管理テーブルを参照して、削除が指示された前記検索条件式に含まれる前記検索タームに対応する該検索条件式を指定したユーザのユーザ識別子を前記ユーザリストから削除するユーザリスト削除ステップと、

削除が指示された前記検索条件式に関連するユーザの管理情報を、前記検索ターム数カウントテーブルから削除する検索ターム数カウントテーブル削除ステップを有することを特徴とする文書検索配送方法。

【請求項 11】 一つ以上の情報源から入手した文書情報のテキストデータを対象として、

1 個以上の検索タームを含む 1 人以上のユーザが指定した検索条件式を登録する検索条件式登録手段と、

テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布ステップを有する文書検索配送装置において、

前記テキスト検索配布手段は、前記テキストをただ一度走査することによって前記複数の検索条件式の該テキストに対する成否を判断するテキスト検索手段を有することを特徴とする文書検索配送装置。

【請求項 12】 請求項 11 記載の文書検索配送装置において、

前記検索条件式登録手段は、

前記検索条件式から全ての検索タームを抽出する検索条件式解析手段と、

ユーザ毎にユーザと該ユーザの検索条件式から抽出された全ての検索タームの数を含む管理情報を格納する検索ターム数カウントテーブルを作成する検索ターム数カウントテーブル作成手段と、

前記検索条件式から抽出した全ての検索タームを、テキストのただ一度の走査により照合する際に参照する多重文字列照合テーブルを生成する多重文字列照合テーブル生成手段と、

検索条件式から抽出された各検索ターム対応に該検索条件式を指定したユーザのユーザ識別子をリストとしてつないだユーザリストを生成するユーザリスト生成手段を有し、

前記テキスト検索配布手段は、

該テキストに対する前記検索条件式の成否の判断時に、前記多重文字列照合テーブルを参照して該テキストを走査することによって、前記検索条件式解析手段により抽出された全ての検索タームを照合するテキスト走査手段と、

前記テキスト走査手段によって照合された検索タームと前記ユーザリストと前記検索ターム数カウントテーブルを照合することにより、該テキストに対する前記検索条件式の成否を判断する検索条件式成否判断手段を有することを特徴とする文書検索配送装置。

【請求項 13】 請求項 12 記載の文書検索配送装置において、

前記多重文字列照合テーブルとして有限オートマトンを用いることを特徴とする文書検索配送装置。

【請求項 14】 請求項 12 または請求項 13 記載の文書検索配送装置において、

前記検索条件式成否判断手段は、

前記ユーザリストを参照し、前記テキスト走査手段によって照合された検索タームの個数をユーザ毎に算出する検索ターム照合数算出手段と、

前記検索ターム照合数算出手段で算出された検索タームの個数と前記検索ターム数カウントテーブルに格納され

ている検索タームの個数とを比較し、一致している場合には該検索タームが含まれる検索条件式が成立しているものとみなす検索ターム数比較手段を有することを特徴とする文書検索配送装置。

【請求項 15】 一つ以上の情報源から入手した文書情報のテキストデータを対象として、

1 個以上の検索タームを含む 1 人以上のユーザが指定した検索条件式を登録する検索条件式登録手段と、  
テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布手段を有する文書検索配送装置において、

1 人以上のユーザあるいはシステム管理者が指定したテキスト配布の条件を記した配布条件を含む配布条件設定式を登録する配布条件設定式登録手段を有し、

前記テキスト検索配布手段は、

前記テキストをただ一度走査することによって前記複数の検索条件式の該テキストに対する成否を判断するテキスト検索手段と、

前記テキスト検索手段によって前記検索条件式が成立したユーザに対して、前記配布条件設定式登録手段によって登録された前記配布条件が成立した時点で前記テキストを配布するテキスト配布制御手段を有することを特徴とする文書検索配送装置。

【請求項 16】 請求項 15 記載の文書検索配送装置において、

前記配布条件設定式登録手段は、

前記配布条件設定式から配布条件を設定すべきユーザの識別子と配布条件を抽出する配布条件設定式解析手段と、

前記配布条件設定式解析手段において前記配布条件設定式から抽出されたユーザの識別子と配布条件を格納した配布条件管理テーブルを作成する配布条件管理テーブル作成手段を有し、

前記テキスト配布制御手段は、

前記配布条件管理テーブルを参照して前記配布条件の成否を判断する配布条件成否判断手段と、

前記配布条件成否判断手段によって前記配布条件が成立していると判断された時点でユーザに対して前記テキストを配布するテキスト配布手段を有することを特徴とする文書検索配送装置。

【請求項 17】 請求項 16 記載の文書検索配送装置において、

前記配布条件として、配布する時間、配布する件数またはテキスト検索から配布までの遅延時間を用いることを特徴とする文書検索配送装置。

【請求項 18】 一つ以上の情報源から入手した文書情報のテキストデータを対象として、

1 個以上の検索タームを含む 1 人以上のユーザが指定した検索条件式を登録する検索条件式登録手段と、

テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布手段を有する文書検索配送装置において、

前記検索条件式の削除が指示された場合には該検索条件式を削除する検索条件式削除手段を有することを特徴とした文書検索配送装置。

【請求項 19】 請求項 18 記載の文書検索配送装置において、

前記検索条件式登録手段は、

前記検索条件式から全ての検索タームを抽出する検索条件式解析手段と、

ユーザ毎にユーザと該ユーザの検索条件式から抽出された全ての検索タームの数を含む管理情報を格納する検索ターム数カウントテーブルを作成する検索ターム数カウントテーブル作成手段と、

前記検索条件式から抽出した全ての検索タームを、テキストのただ一度の走査により照合する際に参照する多重文字列照合テーブルを生成する多重文字列照合テーブル生成手段と、

検索条件式から抽出された各検索ターム対応に該検索条件式を指定したユーザのユーザ識別子をリストとしてつないだユーザリストを生成するユーザリスト生成手段を有し、

前記検索条件式削除手段は、

削除が指示された前記検索条件式に関連する情報を前記検索ターム数カウントテーブルおよび前記ユーザリストから削除する検索条件式管理テーブル削除手段を有することを特徴とする文書検索配送装置。

【請求項 20】 請求項 19 記載の文書検索配送装置において、

前記検索条件式登録手段は、

さらに、前記検索条件式解析手段により抽出された検索タームを格納した検索ターム管理テーブルを作成する検索ターム管理テーブル作成手段を有し、

前記検索条件式管理テーブル削除手段は、

前記検索ターム管理テーブルを参照して、削除が指示された前記検索条件式に含まれる前記検索タームに対応する該検索条件式を指定したユーザのユーザ識別子を前記ユーザリストから削除するユーザリスト削除手段と、

削除が指示された前記検索条件式に関連するユーザの管理情報を、前記検索ターム数カウントテーブルから削除する検索ターム数カウントテーブル削除手段を有することを特徴とする文書検索配送装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、電子メールや情報収集ロボット等を用いて通信社や新聞社等の複数の情報源から入手した電子化文書を、ユーザが予め登録しておいた検索条件式で検索し、条件が成立したユーザに対し

てその電子化文書を配布する文書検索配送システムに係わり、特にユーザ数が増えても電子化文書を一度走査するだけで全てのユーザの検索を終了できる即時性の高いテキスト検索配布機能を有する文書検索配送システムに関する。

#### 【0002】

【従来の技術】近年、電子メールや電子ニュース等により、大量の電子化文書（以下、テキストと呼ぶ）が時々刻々とユーザのもとへ送られてくるようになった。また、インターネットを介して情報を提供する情報源が急激に増えており、これらの情報源から情報収集ロボット等を用いて収集するテキストも膨大な量となっている。このため、これらのテキストを検索し、そのテキストを求めているユーザに即座に配布する文書検索配送システムへのニーズが高まってきている。この文書検索配送システムを実現するための核として、文書検索が用いられる。この文書検索については、“Efficient String Matching: An Aid to Bibliographic Search” (Alfred V. Aho and Margaret J. Corasick, Communications of the ACM, June 1975, Vol. 18, No. 6, pp. 333-340)（以下、従来技術1と呼ぶ）に具体的に記載されている。これは、複数の照合すべき検索文字列（以下、検索タームと呼ぶ）からパターンマッチングマシンと呼ばれる一種の有限オートマトンを作成し、これにより、テキストをただ一度走査するだけで、同時に複数の検索タームを照合することができる方式である。

【0003】本方式について図2を用いて説明する。同図は、“he”、“she”、“his”および“hers”という4つの検索タームを照合する有限オートマトンの状態遷移図である。ここで、円形は有限オートマトンの状態を、実線の矢印は状態遷移を表している。各実線の矢印に付記されたアルファベットはこれに対応した状態遷移が起きる入力文字を、各円形の内部に記された数値は同状態の状態番号を示す。また、破線の矢印はこの有限オートマトンに示されていない文字が入力された場合（以下、フェイルと呼ぶ）の遷移先を示している。ここで、状態1、2、3、6、8から状態0への破線矢印は省略してある。このフェイルによる遷移先は、実際には、図3に示すようなフェイル先状態番号テーブルによって管理される。また、フェイルによって遷移した場合には、該入力文字に対し遷移先の状態で再照合を行う。テキスト走査時に状態2、5、7、9に到達した場合、検索タームと一致する部分文字列がテキスト中に現われたことになるが、これは図4に示すような出力テーブルを参照することによって検出される。この出力テーブルには、状態番号とその状態に到達したときに出力される文字列、すなわちテキスト中の部分文字列と一致した検索タームが格納されている。以下、本方式の動作について、図2を用いて説明する。初期状態は状態0である。この例の場合、入力文字が“h”ならば状態1へ

遷移し、“s”ならば状態3へ遷移する。もし、ここで、これら以外の文字（→[h, s]で表し、“→”は次の要素に否定条件が掛かることを示す）が入ってきた場合は初期状態である状態0に戻る。また、状態3で入力文字が“h”ならば状態4へ遷移する。もし、ここで“h”以外の文字が入ってきた場合は状態0へ戻り、ここで再度照合処理を行う。一方、状態4で入力文字が“e”ならば状態5へ遷移し、図4の出力テーブルを参照することによって検索ターム“she”および“he”と一致する部分文字列がテキスト中に現われたことが検出される。ここで、もし“e”以外の例えば“i”が入力されたときは、フェイルの破線矢印を参照して状態1へ遷移する。そして、遷移先の状態1で該入力文字“i”について再照合することにより状態6へ遷移する。

【0004】次に、テキスト「ushers」を対象に検索タームの照合を行った場合について説明する。まず、第一文字目“u”が入力される。しかし、“u”は“h”および“s”以外の文字なので初期状態である状態0に戻る。次に、第二文字目“s”が入力されることにより、状態0から状態3へ遷移する。以下、第三文字目“h”、第四文字目“e”が入力されることにより、状態4、状態5へ遷移し、図4の出力テーブルを参照することにより、検索ターム“she”および“he”と一致する部分文字列がテキスト中に現われたことが検出される。次に、第五文字目“r”が入力される。しかし、状態5には、入力文字“r”に対する遷移先が存在しないためフェイルとなり、状態2へ遷移する。ここで、“r”に対して再照合を行うことにより、状態8へ遷移する。最後に、第六文字目“s”が入力されることにより、状態9へ遷移し、図4の出力テーブルを参照することにより、検索ターム“hers”と一致する部分文字列がテキスト中に現われたことが検出される。このように、従来技術1には、テキストに対するただ一度の走査で、同時に複数の検索タームを照合することができる文書検索方法が開示されている。

【0005】従来技術1に記載された方式を日本語に拡張したものが“日本語テキスト用のAho-Corasick型パターン照合アルゴリズム”（篠原 武、有川 節夫、情報処理学会研究会資料、自然言語処理、1985. 11. 15、Vol. 8 6、No. 48、pp. 52. 4. 1-52. 4. 8）（以下、従来技術2と呼ぶ）に記載されている。日本語は英語と異なり文字種が多い。そのため計算機内部では通常1文字を2バイト、すなわち英語2文字分で日本語の1文字を表現している。この2バイト文字を1バイトに分割して、従来技術1のような有限オートマトンを単純に作成したのでは、2バイト文字の一部である1バイトと英語の1バイト文字との区別がつかないため、ノイズが発生する可能性がある。そこで従来技術2では、日本工業規格で情報交換用の符号系について1バイト文字と2バイト文字の切り

換えを示す3バイトの文字コード(KIおよびKOで表す)を規定していることに着目し、図5のように1バイト文字と2バイト文字を区別する有限オートマトンを作成することによってこの問題を解決している。

【0006】なお、1バイトは8ビットとし、1バイト文字は00(16)からFF(16)までの256文字である。2バイト文字は2バイトから構成され、前半1バイト、後半1バイトはともに21(16)から7E(16)までである。全部で94×94ある。3バイトの文字コードKIは1B2442(16)である。3バイトの文字コードKOは1B284A(16)である。1バイト文字のうち英数字は21(16)から7E(16)までに属しており、カナ文字はA0(16)からDF(16)に属している。

【0007】ここで、状態0から状態0への遷移は、1バイト文字照合用有限オートマトンへ遷移する1バイト文字以外かKI以外の1バイト文字が入力されたときに実行される。また、状態3から状態3あるいは状態6への遷移は、2バイト文字照合用有限オートマトンへ遷移する2バイト文字の上位バイト以外かKO以外の1バイトコードが入力されたときに実行され、その1バイトコードが1バイト文字であったときに状態3へ遷移し、2バイト文字の上位バイトであったときに状態6へ遷移する。ここで、状態6は2バイト文字照合用有限オートマトンへ遷移しない2バイト文字の上位バイトが入力された場合にバイトずれを防ぐために設けられている。

【0008】バイトずれとは、2バイト文字の下位バイトが2バイト文字の上位バイトであるとみなされてしまうことである。本方式では、状態6を設け、2バイト文字の下位バイトが入力されないと状態3へ戻れないようにすることによってこのバイトずれを防いでいる。この有限オートマトンの動作は従来技術1とまったく同様である。このように、従来技術2に記載された方法を用いることにより、日本語のような1バイト文字と2バイト文字が混在する言語を対象とした場合でも、テキストに対するただ一度の走査で、同時に複数の検索タームを照合することが可能となる。

【0009】なお、上記では、テキストから検索タームを探す手法として有限オートマトンの例を説明したが、他の手法として、「5種類のパターン・マッチング手法をC言語の関数で実現する」 NIKKEI BYT E, August 1987, p.p.175-189

(従来技術3)には拡張BM法について記載されている。拡張BM法は、高速なパターン・マッチング手法であるBM(Boyer-moore)法を複数のパターン(検索文字列)が扱えるように拡張したものであり、上記文献ではEBM(Expanded-Boyer-moore)法と呼んでいる。また、「高速な複数文字列照合アルゴリズム:FAST」 情報処理学会論文誌, Vol.30, NO.9, September 1989(従来技術4)には拡張BM法ではないが、多重

文字列照合を行う手法について記載されている。この手法は、高速なパターン・マッチング手法であるBM法と、有限オートマトンを用いて複数パターンを同時に照合するAC(Aho-Corasick)法の基本的なアイディアを組み合わせたものである。

【0010】

【発明が解決しようとする課題】以上説明した二つの従来技術に示された文書検索方法によると、テキストをただ一度走査するだけで、同時に複数の検索タームを照合することが可能となる。しかし、多数のユーザの検索条件式に対してテキスト検索を行う場合には以下に示す問題が生じる。

(1) ユーザ識別の問題

多数のユーザの検索条件式中に含まれる全ての検索タームで1つの有限オートマトンを作成することにより、テキストの一度の走査で全ての検索タームを照合することが可能となる。しかし、テキスト中の部分文字列と一致した検索タームがどのユーザの検索条件式中に含まれるものであるかを判別できないため、どのユーザの検索条件式が成立したのかが分からない。

(2) 処理時間の問題

各ユーザの検索条件式毎に、その検索条件式中に含まれる検索タームで有限オートマトンを作成すれば、どのユーザの検索条件式が成立しているのかを判別することが可能となる。しかし、有限オートマトンの数分(すなわち、ユーザ数分)だけテキストを走査しなければならないため、ユーザ数が増えるとその分検索に時間が掛かることになる。なお、有限オートマトンに代えて前述の従来技術3、4に示された手法を用いた場合についても同様である。こうした問題に対し、本発明では以下の課題を解決することを目的とする。すなわち、本発明の目的は、複数の情報源から入手したテキストを、ユーザが予め登録しておいた検索条件式に基づき、テキストのただ一度の走査で複数のユーザの検索条件式が成立しているかどうかを判別し、条件が成立しているユーザに対してそのテキストを配布する文書検索配送システムを提供することにある。

【0011】

【課題を解決するための手段】上記目的を達成するため、本発明は、一つ以上の情報源から入手した文書情報のテキストデータを対象として、1個以上の検索タームを含む1人以上のユーザが指定した検索条件式を登録する検索条件式登録ステップと、テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布ステップを有する文書検索配送方法において、前記テキストをただ一度走査することによって前記複数の検索条件式の該テキストに対する成否を判断するテキスト検索ステップを有するようにしている。さら

に、前記検索条件式登録ステップは、前記検索条件式から全ての検索タームを抽出する検索条件式解析ステップと、ユーザ毎にユーザと該ユーザの検索条件式から抽出された全ての検索タームの数を含む管理情報を格納する検索ターム数カウントテーブルを作成する検索ターム数カウントテーブル作成ステップと、前記検索条件式から抽出した全ての検索タームを、テキストのただ一度の走査により照合する際に参照する多重文字列照合テーブルを生成する多重文字列照合テーブル生成ステップと、検索条件式から抽出された各検索ターム対応に該検索条件式を指定したユーザのユーザ識別子をリストとしてつないだユーザリストを生成するユーザリスト生成ステップを有し、前記テキスト検索配布ステップは、該テキストに対する前記検索条件式の成否の判断時に、前記多重文字列照合テーブルを参照して該テキストを走査することによって、前記検索条件式解析ステップにより抽出された全ての検索タームを照合するテキスト走査ステップと、前記テキスト走査ステップによって照合された検索タームと前記ユーザリストと前記検索ターム数カウントテーブルを照合することにより、該テキストに対する前記検索条件式の成否を判断する検索条件式成否判断ステップを有するようにしている。さらに、前記多重文字列照合テーブルとして有限オートマトンを用いるようにしている。さらに、前記検索条件式成否判断ステップは、前記ユーザリストを参照し、前記テキスト走査ステップによって照合された検索タームの個数をユーザ毎に算出する検索ターム照合数算出ステップと、前記検索ターム照合数算出ステップで算出された検索タームの個数と前記検索ターム数カウントテーブルに格納されている検索タームの個数とを比較し、一致している場合には該検索タームが含まれる検索条件式が成立しているものとみなす検索ターム数比較ステップを有するようにしている。

【0012】また、一つ以上の情報源から入手した文書情報のテキストデータを対象として、1個以上の検索タームを含む1人以上のユーザが指定した検索条件式を登録する検索条件式登録ステップと、テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布ステップを有する文書検索配送方法において、1人以上のユーザあるいはシステム管理者が指定したテキスト配布の条件を記した配布条件を含む配布条件設定式を登録する配布条件設定式登録ステップを有し、前記テキスト検索配布ステップは、前記テキストをただ一度走査することによって前記複数の検索条件式の該テキストに対する成否を判断するテキスト検索ステップと、前記テキスト検索ステップによって前記検索条件式が成立したユーザに対して、前記配布条件設定式登録ステップによって登録された前記配布条件が成立した時点で前記テキストを配布するテキスト配布制御ステップを有するようにしている。さらに、前記配

布条件設定式登録ステップは、前記配布条件設定式から配布条件を設定すべきユーザの識別子と配布条件を抽出する配布条件設定式解析ステップと、前記配布条件設定式解析ステップにおいて前記配布条件設定式から抽出されたユーザの識別子と配布条件を格納した配布条件管理テーブルを作成する配布条件管理テーブル作成ステップを有し、前記テキスト配布制御ステップは、前記配布条件管理テーブルを参照して前記配布条件の成否を判断する配布条件成否判断ステップと、前記配布条件成否判断ステップによって前記配布条件が成立していると判断された時点でユーザに対して前記テキストを配布するテキスト配布ステップを有するようにしている。さらに、前記配布条件として、配布する時間、配布する件数またはテキスト検索から配布までの遅延時間を用いるようにしている。

【0013】また、一つ以上の情報源から入手した文書情報のテキストデータを対象として、1個以上の検索タームを含む1人以上のユーザが指定した検索条件式を登録する検索条件式登録ステップと、テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布ステップを有する文書検索配送方法において、前記検索条件式の削除が指示された場合には該検索条件式を削除する検索条件式削除ステップを有するようにしている。さらに、前記検索条件式登録ステップは、前記検索条件式から全ての検索タームを抽出する検索条件式解析ステップと、ユーザ毎にユーザと該ユーザの検索条件式から抽出された全ての検索タームの数を含む管理情報を格納する検索ターム数カウントテーブルを作成する検索ターム数カウントテーブル作成ステップと、前記検索条件式から抽出した全ての検索タームを、テキストのただ一度の走査により照合する際に参照する多重文字列照合テーブルを生成する多重文字列照合テーブル生成ステップと、検索条件式から抽出された各検索ターム対応に該検索条件式を指定したユーザのユーザ識別子をリストとしてつないだユーザリストを生成するユーザリスト生成ステップを有し、前記検索条件式削除ステップは、削除が指示された前記検索条件式に関連する情報を前記検索ターム数カウントテーブルおよび前記ユーザリストから削除する検索条件式管理テーブル削除ステップを有するようにしている。さらに、前記検索条件式登録ステップは、さらに、前記検索条件式解析ステップにより抽出された検索タームを格納した検索ターム管理テーブルを作成する検索ターム管理テーブル作成ステップを有し、前記検索条件式管理テーブル削除ステップは、前記検索ターム管理テーブルを参照して、削除が指示された前記検索条件式に含まれる前記検索タームに対応する該検索条件式を指定したユーザのユーザ識別子を前記ユーザリストから削除するユーザリスト削除ステップと、削除が指示された前記検索条件式に

関連するユーザの管理情報を、前記検索ターム数カウントテーブルから削除する検索ターム数カウントテーブル削除ステップを有するようにしている。

【0014】また、一つ以上の情報源から入手した文書情報のテキストデータを対象として、1個以上の検索タームを含む1人以上のユーザが指定した検索条件式を登録する検索条件式登録手段と、テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布手段を有する文書検索配送装置において、前記テキスト検索配布手段は、前記テキストをただ一度走査することによって前記複数の検索条件式の該テキストに対する成否を判断するテキスト検索手段を有するようにしている。さらに、前記検索条件式登録手段は、前記検索条件式から全ての検索タームを抽出する検索条件式解析手段と、ユーザ毎にユーザと該ユーザの検索条件式から抽出された全ての検索タームの数を含む管理情報を格納する検索ターム数カウントテーブルを作成する検索ターム数カウントテーブル作成手段と、前記検索条件式から抽出した全ての検索タームを、テキストのただ一度の走査により照合する際に参照する多重文字列照合テーブルを生成する多重文字列照合テーブル生成手段と、検索条件式から抽出された各検索ターム対応に該検索条件式を指定したユーザのユーザ識別子をリストとしてつないだユーザリストを生成するユーザリスト生成手段を有し、前記テキスト検索配布手段は、該テキストに対する前記検索条件式の成否の判断時に、前記多重文字列照合テーブルを参照して該テキストを走査することによって、前記検索条件式解析手段により抽出された全ての検索タームを照合するテキスト走査手段と、前記テキスト走査手段によって照合された検索タームと前記ユーザリストと前記検索ターム数カウントテーブルを照合することにより、該テキストに対する前記検索条件式の成否を判断する検索条件式成否判断手段を有するようにしている。さらに、前記多重文字列照合テーブルとして有限オートマトンを用いるようにしている。さらに、前記検索条件式成否判断手段は、前記ユーザリストを参照し、前記テキスト走査手段によって照合された検索タームの個数をユーザ毎に算出する検索ターム照合数算出手段と、前記検索ターム照合数算出手段で算出された検索タームの個数と前記検索ターム数カウントテーブルに格納されている検索タームの個数とを比較し、一致している場合には該検索タームが含まれる検索条件式が成立しているものとみなす検索ターム数比較手段を有するようにしている。

【0015】また、一つ以上の情報源から入手した文書情報のテキストデータを対象として、1個以上の検索タームを含む1人以上のユーザが指定した検索条件式を登録する検索条件式登録手段と、テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断

し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布手段を有する文書検索配送装置において、1人以上のユーザあるいはシステム管理者が指定したテキスト配布の条件を記した配布条件を含む配布条件設定式を登録する配布条件設定式登録手段を有し、前記テキスト検索配布手段は、前記テキストをただ一度走査することによって前記複数の検索条件式の該テキストに対する成否を判断するテキスト検索手段と、前記テキスト検索手段によって前記検索条件式が成立したユーザに対して、前記配布条件設定式登録手段によって登録された前記配布条件が成立した時点で前記テキストを配布するテキスト配布制御手段を有するようにしている。さらに、前記配布条件設定式登録手段は、前記配布条件設定式から配布条件を設定すべきユーザの識別子と配布条件を抽出する配布条件設定式解析手段と、前記配布条件設定式解析手段において前記配布条件設定式から抽出されたユーザの識別子と配布条件を格納した配布条件管理テーブルを作成する配布条件管理テーブル作成手段を有し、前記テキスト配布制御手段は、前記配布条件管理テーブルを参照して前記配布条件の成否を判断する配布条件成否判断手段と、前記配布条件成否判断手段によって前記配布条件が成立していると判断された時点でユーザに対して前記テキストを配布するテキスト配布手段を有するようにしている。さらに、前記配布条件として、配布する時間、配布する件数またはテキスト検索から配布までの遅延時間を用いるようにしている。

【0016】また、一つ以上の情報源から入手した文書情報のテキストデータを対象として、1個以上の検索タームを含む1人以上のユーザが指定した検索条件式を登録する検索条件式登録手段と、テキストを入手した際に、該テキストに対する前記検索条件式の成否を判断し、該検索条件式が成立したユーザに対して、該テキストを配布するテキスト検索配布手段を有する文書検索配送装置において、前記検索条件式の削除が指示された場合には該検索条件式を削除する検索条件式削除手段を有するようにしている。さらに、前記検索条件式登録手段は、前記検索条件式から全ての検索タームを抽出する検索条件式解析手段と、ユーザ毎にユーザと該ユーザの検索条件式から抽出された全ての検索タームの数を含む管理情報を格納する検索ターム数カウントテーブルを作成する検索ターム数カウントテーブル作成手段と、前記検索条件式から抽出した全ての検索タームを、テキストのただ一度の走査により照合する際に参照する多重文字列照合テーブルを生成する多重文字列照合テーブル生成手段と、検索条件式から抽出された各検索ターム対応に該検索条件式を指定したユーザのユーザ識別子をリストとしてつないだユーザリストを生成するユーザリスト生成手段を有し、前記検索条件式削除手段は、削除が指示された前記検索条件式に関連する情報を前記検索ターム数カウントテーブルおよび前記ユーザリストから削除する



検索条件式管理テーブル削除手段を有するようにしている。さらに、前記検索条件式登録手段は、さらに、前記検索条件式解析手段により抽出された検索タームを格納した検索ターム管理テーブルを作成する検索ターム管理テーブル作成手段を有し、前記検索条件式管理テーブル削除手段は、前記検索ターム管理テーブルを参照して、削除が指示された前記検索条件式に含まれる前記検索タームに対応する該検索条件式を指定したユーザのユーザ識別子を前記ユーザリストから削除するユーザリスト削除手段と、削除が指示された前記検索条件式に関連するユーザの管理情報を、前記検索ターム数カウントテーブルから削除する検索ターム数カウントテーブル削除手段を有するようにしている。

【0017】

【発明の実施の形態】以下に、本発明の実施例を図を参照して説明する。

《第一実施例》最初に、第一実施例の概略説明を図6を参照して行う。まず、検索条件式登録処理について説明する。まず、検索条件式を解析し、検索条件式中に含まれる検索タームを抽出する。そして、抽出された検索タームの数を、検索ターム数カウントテーブル作成処理により検索ターム数カウントテーブルに格納する。例えば、図6において、ユーザ1：「“文書”と“検索”が含まれる文書」という検索条件式には“文書”と“検索”という2つの検索タームが含まれているので、検索ターム数カウントテーブルのユーザ1に対応する箇所に2を格納する。同様に、ユーザ2、ユーザ3に対応する箇所に1、2をそれぞれ格納する。次に、有限オートマトン作成処理で、上記検索条件式解析で抽出された全ての検索タームを照合する有限オートマトンを作成する。この有限オートマトンは、従来技術1および従来技術2に示されたものと同様である。

ユーザ1：「“文書”と“検索”が含まれる文書」、ユーザ2：「“登山”が含まれる文書」およびユーザ3：「“検索”と“登録”が含まれる文書」という検索条件式から抽出した検索ターム“文書”、“検索”、“登山”および“登録”を照合する有限オートマトンの状態遷移図は図に示されたものとなる。本図では、簡単のため、状態遷移を2バイト（1文字）単位で示す。次に、ユーザリスト作成処理で、それぞれの検索タームを指定したユーザの識別子をユーザリストとして、有限オートマトンに接続する。図6では、例えば“検索”を照合するとその末尾の状態4からユーザリストが参照され、“検索”を指定したユーザが“ユーザ1”および“ユーザ3”であることが検出される。

【0018】次に、テキストの検索および配布処理であるテキスト検索配布処理について説明する。この処理では、まずテキスト走査処理でテキストの走査を行い、検索タームを照合する。例えば、テキスト：「文書を検索する」を図6に示した有限オートマトンを用いて走査し

た場合には、“文書”および“検索”と一致する部分文字列がテキスト中に現われたことが検出される。本図に示した有限オートマトンで、末尾の状態に“○”が記されている検索タームはテキスト中に一致する部分文字列が出現したことを示し、“×”が記されている検索タームはテキスト中に一致する部分文字列が出現しなかったことを示す。本例では、“文書”および“検索”と一致する部分文字列がテキスト中に現われたので、その末尾の状態である状態2および状態4に“○”が記されている。

【0019】次に、検索ターム数カウント処理でこれらテキスト中の部分文字列と一致した検索タームの出現数をユーザ毎にカウントする。例えば、ユーザ1に対しては“文書”および“検索”が一致しているので2と、ユーザ3に対しては“検索”だけが一致しているので1とカウントする。しかし、ユーザ2はテキスト中に検索タームと一致する部分文字列が現われなかったので0である。最後に、検索条件式チェック処理で、上記検索ターム数カウントテーブルに格納された検索ターム数と上記検索ターム数カウント処理で算出した検索ターム出現カウント数とを比較し、一致している場合には、テキスト配布処理でそのユーザに対しテキストを配布する。例えば、図6でユーザ1は検索ターム数が2で一致しているため、テキストを配布するが、ユーザ2およびユーザ3は一致していないので配布しない。

【0020】以上のように、本実施例では、有限オートマトンを用いてテキストを走査し、テキスト中に一致する部分文字列として現われた検索タームの出現数を、ユーザリストを参照しながら各ユーザ毎にカウントする。そして、カウントした結果と検索ターム数カウントテーブルに予め格納しておいた検索ターム数とを比較することによって検索条件式が成立しているかどうかをチェックする。この結果、テキストのただ一度の走査で複数のユーザの検索条件式が成立しているかどうかを判別することが可能となり、即時性の高いテキスト検索配布が実現できることになる。

【0021】以下、本発明の第一の実施例について図1を用いて詳細に説明する。本発明を適用した文書検索配送システムは、ディスプレイ100、キーボード101、中央演算処理装置（CPU）102、主メモリ104およびこれらを結ぶバス103から構成される。また、バス103には、LAN（Local Area Network）等の通信回線124を介して、ニュースを配信するニュース配信元125や文書検索配送システムを利用するユーザ126が接続されている。ニュース配信元125は電子メールや電子ニュース等を用いてニュースデータを電子化したテキストを本システムへ配信したり、インターネットを介してテキストを提示し、ユーザ126は電子メールを用いて検索条件式を本システムへ登録する。本システムからは上記検索条件式に基づいて検索された上

記テキストが電子メールを用いて該当ユーザへ配布される。以下、本実施例では、ニュース配信元125は電子メール等を用いて本システムにテキストを配信するものとして述べるが、ニュース配信元125はテキストをインターネット上に提示するだけで、テキストの収集は情報収集ロボットを用いて行うようにしてもかまわない。

【0022】主メモリ104には、システム制御プログラム105、検索条件式登録制御プログラム106、検索条件式解析プログラム107、検索ターム数カウントテーブル作成プログラム108、検索用オートマトン作成プログラム109、テキスト検索制御プログラム112、テキスト取得プログラム113、テキスト検索プログラム114、テキスト成形プログラム118、電子メールプログラム119、検索ターム数カウントテーブル120、有限オートマトン121およびユーザリスト122が格納されるとともにワークエリア123が確保される。検索用オートマトン作成プログラム109は有限オートマトン作成プログラム110およびユーザリスト作成プログラム111で構成される。なお、本実施例では、テキストから検索タームを探し出す手法として有限オートマトンを用いているが、テキストから検索タームを探し出す手法としては有限オートマトンに限られるものではなく、有限オートマトンを用いてもよいし、前述した従来技術3、4に示された手法を用いてもよく、この場合、検索用オートマトン作成プログラム、有限オートマトン作成プログラムの表現は適当ではなく、より一般化した表現を用いると、検索用文字列照合テーブル作成プログラム、多重文字列照合テーブル作成プログラムの様になる。また、テキスト検索プログラム114はテキスト走査プログラム115、検索ターム数カウントプログラム116および検索条件式チェックプログラム117で構成される。以上のプログラムはハードディスク装置（本図には示していない）、フロッピーディスク

（本図には示していない）などのコンピュータで読み書きできる記憶媒体に格納することもできる。

【0023】システム制御プログラム105は文書検索配送システムの管理者によるキーボード101からの指示を受け起動する。検索条件式登録制御プログラム106およびテキスト検索制御プログラム112はユーザ126からの検索条件式の登録指示やニュース配信元125からのテキストの配信により、システム制御プログラム105によって起動され、それぞれ検索条件式解析プログラム107、検索ターム数カウントテーブル作成プログラム108および検索用オートマトン作成プログラム109の制御と、テキスト取得プログラム113、テキスト検索プログラム114およびテキスト成形プログラム118の制御を行う。電子メールプログラム119にはワークステーションなどで一般的に用いられている既存のメールプログラムを用いる。本電子メールプログラム119はテキスト検索制御プログラム112の処理

結果に応じて、システム制御プログラム105によって起動される。

【0024】以下、本実施例における文書検索配送システムの処理内容について説明する。まず、システム制御プログラム105の処理内容について図7のPAD（Problem Analysis Diagram）図を用いて説明する。システム制御プログラム105では、まずステップ700で、キーボード101から終了コマンドが入力されるまで、以下のステップを繰り返す。この繰り返し処理では、まずステップ701でユーザ126から電子メールによって検索条件式が送られてきているかどうかを調べる。ここで、検索条件式が送られてきている場合には、ステップ702で検索条件式登録制御プログラム106を起動して、検索条件式の登録を行う。次に、ステップ703でニュース配信元125から電子メールによってテキストが送られてきているかどうかを調べる。ここで、テキストが送られてきている場合には、ステップ704でテキスト検索制御プログラム112を起動して、テキストの検索を行う。次に、ステップ705でテキスト検索制御プログラム112におけるテキスト検索の結果を調べ、成立している検索条件式が一つでも存在すると判断された場合には、ステップ706で電子メールプログラム119を起動し、成立した検索条件式を指定したユーザに対し、該当テキストを電子メールを用いて配布する。以上がシステム制御プログラム105の処理内容である。

【0025】次に、検索条件式登録制御プログラム106による検索条件式登録の処理内容について図8のPAD図を用いて説明する。検索条件式登録制御プログラム106はシステム制御プログラム105によって起動される。本プログラムは、まずステップ800で検索条件式解析プログラム107を起動し、ユーザ126から電子メールによって送られてきた検索条件式を解析する。この検索条件式の解析処理では、検索条件式を以下のいずれかの形式に展開する。すなわち、（a）単一検索タームのみ、（b）複数の（a）の論理積条件、（c）複数の（a）と複数の（b）の論理和条件である。つまり、論理和条件の外側に論理積条件が掛からないように検索条件式を展開する。ただし、否定条件は検索条件式全体または検索タームに掛かるようにする。ここで、論理積条件とは、例えば、

検索式「“文書” and “検索”」

のように、「“文書”と“検索”の両方の文字列が現れる文書を探せ」という意味を持ち、論理和条件とは、例えば、

検索式「“文書” or “検索”」

のように、「“文書”か“検索”のどちらかの文字列が現れる文書を探せ」という意味を持つ。また、否定条件とは、例えば、

検索式「¬ “検索”」

のように、「“検索”が現れない文書を探せ」という意味を持つ。例えば、“A”、“B”、“C”、“D”、“E”を検索タームとすると、それぞれ次のような形式に展開される。

- (a) A
- (b) A and B and C and . . .
- (c) (A and B) or C or (D and E) or . . .

例えば、検索条件式が論理和条件に対する論理積条件である場合、すなわち論理和条件の外側に論理積条件が掛かっている場合には、図9に示すように展開して上記の条件を満たすように変形する。ここで、展開結果における論理積条件あるいは単一の検索タームの部分を項と呼ぶ。

【0026】次に、ステップ801で検索条件式登録制御プログラム106は、検索ターム数カウントテーブル作成プログラム108を起動して、検索条件式解析プログラム107の解析結果として得られた検索条件式に含まれる検索タームの数や検索条件式に否定条件が掛かっているかどうかを示す情報を検索ターム数カウントテーブル120に格納する。この検索ターム数カウントテーブル120は、検索条件式毎（すなわちユーザ毎）に、検索条件式中の項に対応してその中に含まれる検索タームの数を格納したテーブルである。図10にその構造を示す。本図に示す検索ターム数カウントテーブル120は、ユーザ番号1：「（“文書” and “検索”） or （“文書” and “サーチ”）」、ユーザ番号2：「“文字” and “認識” and “学習”」およびユーザ番号3：「“（検索” and “学習”）」という3つの検索条件式に対して作成されたものである。まず、この検索ターム数カウントテーブル120の先頭の要素として検索条件式否定フラグを設ける。この検索条件式否定フラグには、検索条件式全体に否定条件が掛かっている場合には1を、そうでなければ0を設定する。例えば、本図に示す例の場合、ユーザ番号3の検索条件式全体に否定条件が掛かっているため1を設定するが、その他の検索条件式には否定条件が掛かっていないので0を設定する。この検索条件式否定フラグに対し第1項から順にその項に含まれる検索タームの数をリストとしてつないでいく。例えば、ユーザ番号1に対応するリストの2番目の要素には、検索条件式の第1項に含まれる検索タームの数が格納されることになるが、本項には“文書”と“検索”という2つの検索タームが含まれているので2を格納する。さらにその次の要素には、第2項に含まれる検索タームの数である2を格納する。また、2番目以降の要素にはそれぞれ、検索タームの数を格納すると同時に、テキスト検索時にテキスト中の部分文字列と一致した検索タームの出現数をカウントするための領域が確保されている。

【0027】最後に、ステップ802で検索条件式登録制御プログラム106は、検索用オートマトン作成プロ

グラム109を起動し、上記検索条件式解析プログラム107の解析結果として得られた検索条件式に含まれる全ての検索タームを照合する有限オートマトン121を作成する。そして、それらの検索タームが含まれる検索条件式を指定したユーザ126の識別子をリストとしてつないだユーザリスト122を作成し、これを有限オートマトン121に接続する。この検索用オートマトン作成プログラム109の処理内容については、後で詳細に説明する。以上が検索条件式登録制御プログラム106による検索条件式登録の処理内容である。

【0028】次に、テキスト検索制御プログラム112によるテキスト検索の処理内容について図11のPAD図を用いて説明する。テキスト検索制御プログラム112はシステム制御プログラム105によって起動される。本プログラムは、まずステップ1100でテキスト取得プログラム113を起動し、ニュース配信元125から電子メール等によって送られてきたテキストをワークエリア123に格納する。次に、ステップ1101でテキスト検索プログラム114を起動し、前記検索ターム数カウントテーブル作成プログラム108によって作成された検索ターム数カウントテーブル120および前記検索用オートマトン作成プログラム109によって作成された有限オートマトン121とユーザリスト122を用いて、ワークエリア123に格納されているテキストを検索する。このテキスト検索プログラム114の処理内容については、後で詳細に説明する。次に、ステップ1102で、テキスト検索プログラム114におけるテキスト検索処理の結果を調べ、成立している検索条件式が一つでも存在した場合には、ステップ1103でテキスト成形プログラム118を起動し、ワークエリア123に格納されているテキストを電子メールプログラム119が配布できるような形式に成形する。以上がテキスト検索制御プログラム112によるテキスト検索の処理内容である。

【0029】次に、図8に示した検索条件式登録制御プログラム106による検索条件式登録処理における検索用オートマトン作成プログラム109の処理内容について、図12のPAD図を用いて説明する。検索用オートマトン作成プログラム109は、図12に示すように、まずステップ1200で有限オートマトン作成プログラム110を起動し、前記検索条件式解析プログラム107によって抽出された全ての検索タームを照合する有限オートマトン121を作成する。この有限オートマトン121の作成方法には、従来技術1および従来技術2に開示されている方法を用いる。次に、ステップ1201でユーザリスト作成プログラム111を起動し、前記検索条件式解析プログラム107によって抽出された検索タームが含まれる検索条件式を指定したユーザ126の識別番号（ユーザ番号）とその検索タームが含まれる項の番号（項番号）からユーザリスト122を作成し、ス

テップ1202でポインタを介して有限オートマトン121の出力テーブルに接続する。

【0030】前記検索ターム数カウントテーブル120の説明に用いた3つの検索条件式から作成される有限オートマトン121とユーザリスト122の例を図13に示す。本図に示した有限オートマトン121は、検索条件式中に含まれる“文書”、“文字”、“検索”、“サーチ”、“認識”および“学習”という6つの検索タームを照合するものである。本図では簡単のため、状態遷移を2バイト（1文字）単位で示している。この有限オートマトン121は従来技術1および従来技術2で示したものと同様なものであるが、出力テーブルの部分が異なる。出力テーブルに格納されている各状態番号に対応して出現フラグが設けられている。この出現フラグはテキストの走査開始時に0にリセットしておき、検索タームと一致する部分文字列がテキスト中に現われた場合に1をセットする。また、出力テーブルの末尾にはポインタが設けられており、その検索タームを含む検索条件式のユーザ番号および項番号をリストとしてつないだユーザリスト122を指し示すようにしている。ユーザリスト122の一つ一つの要素には、検索ターム否定フラグが設けられており、検索条件式においてその検索タームに否定条件が掛かっている場合には1を、そうでなければ0を設定する。例えば、本図において“文書”という検索タームはユーザ番号1の検索条件式の項番号1および2の項に否定条件無しで存在し、“認識”という検索タームはユーザ番号2の検索条件式の項番号1の項に否定条件付きで存在している。

【0031】次に、図11に示したテキスト検索制御プログラム112におけるテキスト検索処理を実行するテキスト検索プログラム114の処理内容について、図14のPAD図を用いて説明する。テキスト検索プログラム114は、本図に示すように、まずステップ1400でテキスト検索の初期設定として、検索ターム数カウントテーブル120に設けられている検索タームの出現数カウント用領域および有限オートマトン121の出力テーブルに設けられている出現フラグを0にリセットする。次に、ステップ1401で、テキスト走査プログラム115を起動し、ワークエリア123に格納されているテキストを、有限オートマトン作成プログラム110によって作成された有限オートマトン121で走査し、検索タームを照合する。この有限オートマトン121による検索タームの照合処理には従来技術1および従来技術2に開示されている方法を用いる。このとき、テキスト中に一致する部分文字列が出現した検索タームについては、その検索タームに対応する出力テーブルの出現フラグを1に設定する。次に、ステップ1402で検索ターム数カウントプログラム116を起動し、テキスト中に一致する部分文字列が出現した検索タームのカウントを行う。これは、出力テーブルに設けられている出現フラグが1である

ユーザリスト122を辿り、検索ターム否定フラグが0であるユーザ番号と項番号に対応する検索ターム数カウントテーブル120における検索タームの出現数カウント用領域の値を1ずつ増やしていくことにより実現する。次に、ステップ1403でテキスト中に一致する部分文字列が出現しなかった検索タームのカウントを行う。これは、出力テーブルに設けられている出現フラグが0であるユーザリスト122を辿り、検索ターム否定フラグが1であるユーザ番号と項番号に対応する検索ターム数カウントテーブル120における検索タームの出現数カウント用領域の値を1ずつ増やしていくことにより実現する。

【0032】次に、ステップ1404で検索条件式チェックプログラム117を起動し、検索ターム数カウントテーブル120を参照し、検索条件式が成立しているかどうかを調べる。ここで、以下の2つの条件のうち、どちらかを満たしている検索条件式は成立しているものとみなせる。

条件（1）：検索条件式否定フラグが0（すなわち、検索条件式に否定条件が掛かっていない）で、検索ターム数が一致している項番号が少なくとも一つある。

【0033】条件（2）：検索条件式否定フラグが1（すなわち、検索条件式に否定条件が掛かっている）で、検索ターム数が一致している項番号が一つもない。

【0034】この検索条件式の成否の判定について、図10を用いて説明する。本発明では本図に示すように、ユーザ126が指定した検索条件式を項が論理和条件でつながれた形式に変形し、その項毎に含まれる検索ターム数を検索ターム数カウントテーブル120に格納している。項が論理和条件でつながれているということは、それらの項のどれか一つが成立すればその検索条件式全体が成立することになる。ここで、項は単一の検索タームあるいは検索タームの論理積条件である。そのため、その項の中に含まれる検索タームと一致する部分文字列全てがテキスト中に出現した場合、すなわち予め検索ターム数カウントテーブル120に格納しておいた検索ターム数と検索ターム数カウントプログラム116によって算出された検索ターム出現カウント数が一致した場合、その項が成立することになる。その結果、項の論理和条件で構成される検索条件式も成立することになる。このように、上記条件（1）を満たせば、検索条件式は成立しているものとみなせる。

【0035】上記条件（2）では、条件（1）と逆になる。検索条件式に否定条件が掛かっている場合、検索条件式から否定条件を取った検索条件式が成立していれば、否定条件が掛かった元の検索条件式は成立せず、検索条件式から否定条件を取った検索条件式が成立していなければ、否定条件が掛かった元の検索条件式は成立していることになる。予め検索ターム数カウントテーブル120に格納しておいた検索ターム数と検索ターム数カ

ウントプログラム 116 によって算出された検索ターム出現カウント数が一致している項が一つもなければ、否定条件を取った検索条件式は成立せず、否定条件が掛かった元の検索条件式が成立することになる。このように、上記条件 (2) を満たせば、検索条件式は成立しているものとみなせる。上記条件のどちらかを満たしている検索条件式は成立しているとみなせるため、ステップ 1405 でそのユーザ番号をテキスト検索制御プログラム 112 へ出力し、本プログラムを終了する。以上が本発明の文書検索方法の実施例である。

【0036】以下、図 8 に示した本実施例における検索条件式登録制御プログラム 106 の処理手順について具体的に説明する。まず、図 8 の検索条件式登録制御プログラム 106 のステップ 800 における検索条件式解析

プログラム 107 の処理について説明する。検索条件式解析プログラム 107 は検索条件式登録制御プログラム 106 によって起動される。例えば、「“文書” and ( “検索” or “サーチ” )」、すなわち「“文書” を含み、かつ “検索” または “サーチ” を含む文書」というユーザ 1 の検索条件式の展開結果を数 (1) に、「“文字” and ¬ ( “認識” or “学習” )」、すなわち「“文字” を含むが、“認識” も “学習” も含まない文書」というユーザ 2 の検索条件式の展開結果を数 (2) に、「¬ ( “検索” and “学習” )」、すなわち「“検索” も “学習” も含まない文書」というユーザ 3 の検索条件式の展開結果を数 (3) に示す。

【0037】

【数 1】

【数 1】

ユーザ 1 : 「“文書” and ( “検索” or “サーチ” )」  
→ 「( “文書” and “検索” ) or ( “文書” and “サーチ” )」  
... (1)

【0038】

【数 2】

【数 2】

ユーザ 2 : 「“文字” and ¬ ( “認識” or “学習” )」  
→ 「“文字” and ( ¬ “認識” and ¬ “学習” )」  
→ 「“文字” and ¬ “認識” and ¬ “学習”」 ... (2)

【0039】

【数 3】

【数 3】

ユーザ 3 : 「¬ ( “検索” and “学習” )」 ... (3)

【0040】この結果、論理和条件の外側に論理積条件が掛かっていない検索条件式が、すなわち、「( “文書” and “検索” ) or ( “文書” and “サーチ” )」、「“文字” and ¬ “認識” and ¬ “学習”」、「¬ ( “検索” and “学習” )」が得られる。これらの検索条件式中に含まれる検索タームをユーザ番号と項番号という観点から表にまとめると表 1 のようになる。

【0041】

【表 1】

【表 1】

項番号 ユーザ番号	1	2
1	文書 検索	文書 サーチ
2	文字 ¬認識 ¬学習	
¬3	検索 学習	

【0042】ここで、ユーザ番号の前に“¬”が付与されているのは検索条件式全体に否定条件が掛かることを、検索タームの前に“¬”が付与されているのは検索

タームに否定条件が掛かることを示す。例えば、ユーザ番号 1 の検索条件式の項番号 1 には“文書”と“検索”という 2 つの検索タームが、項番号 2 には“文書”と“サーチ”という 2 つの検索タームが含まれ、ユーザ番号 2 の検索条件式の項番号 1 には“文字”、“認識”および“学習”という 3 つの検索タームが含まれ、このうち“認識”と“学習”には否定条件が掛かることを表している。

【0043】次に、図 8 の検索条件式登録制御プログラム 106 のステップ 801 における検索ターム数カウントテーブル作成プログラム 108 の処理について説明する。検索ターム数カウントテーブル作成プログラム 108 は検索条件式登録制御プログラム 106 によって、検索条件式解析プログラム 107 の次に起動される。本プログラムは、検索条件式解析プログラム 107 による解析結果に基づき検索ターム数カウントテーブル 120 を作成する。表 1 の解析結果から作成される検索ターム数カウントテーブル 120 は図 10 のようになる。本テーブルには、各ユーザ番号毎に項番号に対応して検索タームの数が格納される。また前述したように、検索条件式否定フラグには検索条件式全体に否定条件が掛かっているなら 1 を、そうでなければ 0 を設定する。

【0044】最後に、図 8 の検索条件式登録制御プログ

ラム 106 のステップ 802 における検索用オートマトン作成プログラム 109 の処理について説明する。本プログラムの処理内容は図 12 に示した。本プログラムは図 1 に示したように、有限オートマトン作成プログラム 110 およびユーザリスト作成プログラム 111 から構成される。以下、順に説明する。

【0045】有限オートマトン作成プログラム 110 では、検索条件式解析プログラム 107 によって抽出された全ての検索タームを照合する有限オートマトン 121 を作成する。例えば、数(1)、数(2)、数(3) および表 1 に示した例の場合、検索条件式解析プログラム 107 の解析結果から“文書”、“検索”、“サーチ”、“文字”、“認識”および“学習”の 6 つの検索タームが得られる。従来技術 1 および従来技術 2 に開示されている方法を用いて、これらの検索タームを照合する有限オートマトン 121 を作成すると、図 13 に示したような有限オートマトン 121 が得られる。ただし、ここでは簡単のため、状態遷移を 1 文字すなわち 2 バイト単位で示している。

【0046】ユーザリスト作成プログラム 111 では、検索条件式解析プログラム 107 によって得られた検索タームが含まれる検索条件式のユーザ番号、項番号およびその検索タームに否定条件が掛かっているかどうかという情報をリストでつなぐことによってユーザリスト 122 を作成し、ポインタを介して有限オートマトン 121 の出力テーブルに接続する。このユーザリスト 122 の作成方法は前述した通りである。表 1 の解析結果から作成されるユーザリスト 122 を図 15 に示す。例えば、“学習”という検索タームはユーザ番号 2 の検索条件式の項番号 1 の項に否定条件付きで含まれ、ユーザ番号 3 の検索条件式の項番号 1 の項に否定条件無しで含まれるので、これらに対応する番号をリストでつないだ形でユーザリスト 122 が作成される。そして、このようにして作成されたユーザリスト 122 は、有限オートマトン 121 の出力テーブルにポインタを介して接続される。以上が本実施例における検索条件式登録制御プログラム 106 における検索条件式登録の詳細な手順である。

【0047】以下、図 11 に示した本実施例におけるテキスト検索制御プログラム 112 の処理手順について具体的に説明する。まず、図 11 のテキスト検索制御プログラム 112 のステップ 1100 におけるテキスト取得プログラム 113 の処理について説明する。テキスト取得プログラム 113 はテキスト検索制御プログラム 112 によって起動される。本プログラムでは、電子メールによって配信されたテキストをワークエリア 123 に格納する。本プログラムによって、「検索した文書の書式を解析し、文字列部分を認識する」というテキストがワークエリア 123 に格納されたものとして、以下の説明を行う。

【0048】図 11 のテキスト検索制御プログラム 112 のステップ 1101 におけるテキスト検索プログラム 114 の処理について説明する。本プログラムの処理内容は図 14 に示した。本プログラムは図 1 に示したように、テキスト走査プログラム 115、検索ターム数カウントプログラム 116 および検索条件式チェックプログラム 117 から構成される。以下、順に説明する。まず、上記テキスト走査プログラム 115、検索ターム数カウントプログラム 116 および検索条件式チェックプログラム 117 が実行される前に初期設定が行われる。ここでは、図 10 および図 15 に示すように、検索ターム数カウントテーブル 120 の検索タームの出現数カウント用領域および出力テーブルの出現フラグが 0 にリセットされる。

【0049】テキスト走査プログラム 115 では、ワークエリア 123 に格納されているテキストを、有限オートマトン 121 で走査して、検索タームを照合する。ここで、テキスト中に一致する部分文字列が出現した検索タームに対応する出現フラグを 1 に設定する。例えば、図 16 に示したようにテキスト「検索した文書の書式を解析し、文字列部分を認識する」を走査すると、まず、“検索”がテキスト中に現われる。

【0050】そこで、“検索”に対応する出現フラグを 1 に設定する。以下、“文書”、“文字”および“認識”の順に出現するので、それらの検索タームに対応する出現フラグを 1 に設定する。“サーチ”および“学習”という検索タームについては、テキスト中に一致する部分文字列が現われないので、それらの検索タームに対応する出現フラグは 0 のままである。

【0051】検索ターム数カウントプログラム 116 では、まず、テキスト中に一致する部分文字列が出現した検索タームのカウントを行う。ここでは、テキスト中に一致する部分文字列が出現し、すなわち出現フラグが 1 で、検索タームに否定条件が掛かっていない、すなわち検索ターム否定フラグが 0 であるユーザ番号と項番号に対応する検索ターム数カウントテーブル 120 の検索タームの出現数カウント用領域を 1 増やす。例えば、図 17 の例では、検索ターム“検索”に関してはユーザ番号 3 の項番号 1 の検索ターム否定フラグは 0 なのでカウントするが、検索ターム“認識”についてはユーザ番号 2 の項番号 1 の検索ターム否定フラグが 1 なのでカウントしない。次に、テキスト中に一致する部分文字列が出現しなかった検索タームのカウントを行う。ここでは、テキスト中に一致する部分文字列が出現せず、すなわち出現フラグが 0 で、検索タームに否定条件が掛かっている、すなわち検索ターム否定フラグが 1 であるユーザ番号と項番号に対応する検索ターム数カウントテーブル 120 の検索タームの出現数カウント用領域を 1 増やす。例えば、図 18 の例では、検索ターム“学習”に関してはユーザ番号 2 の項番号 1 の検索ターム否定フラグは 1

なのでカウントするが、ユーザ番号3の項番号1の検索ターム否定フラグは0なのでカウントしない。

【0052】次に、検索条件式チェックプログラム117では、検索ターム数カウントテーブル120を参照し、検索条件式が成立しているかどうかを調べる。ここで、以下の2つの条件のうち、どちらかを満たしている検索条件式は成立しているとみなせるため、その検索条件式を指定したユーザの番号を出力する。

条件(1)：検索条件式否定フラグが0、すなわち検索条件式に否定条件が掛かっておらず、検索ターム数が一致している項番号が少なくとも一つある。

条件(2)：検索条件式否定フラグが1、すなわち検索条件式に否定条件が掛かっていて、検索ターム数が一致している項番号が一つもない。

例えば、図19の例の場合、ユーザ番号1に対応する検索条件式否定フラグが0で、項番号1の検索ターム数と検索ターム出現カウント数が一致しているため、上記条件(1)を満たしている。また、ユーザ番号3に対応する検索条件式否定フラグが1で、検索ターム数と検索ターム出現カウント数が一致している項番号が存在しないため、上記条件(2)を満たしている。しかし、ユーザ番号2では、検索条件式否定フラグが0であるにもかかわらず、検索ターム数と検索ターム出現カウント数が一致している項が存在しないため、上記条件をいずれも満たしてはいない。したがって、ユーザ番号1およびユーザ番号3の検索条件式が成立しているとみなせるため、これらのユーザ番号を出力する。

【0053】最後に、図11のテキスト検索制御プログラム112のステップ1103におけるテキスト成形プログラム118の処理について説明する。テキスト成形プログラム118は、上記テキスト検索プログラム114の結果、ユーザ番号が出力された場合のみ、テキスト検索制御プログラム112によって起動される。本プログラムでは、ワークエリア123に格納されているテキストを電子メールプログラム119が配布できるような形式に成形する。例えば、テキストの先頭にヘッダと呼ばれる制御情報を付加する。図20に本プログラムの処理結果の例を示す。本図では、ヘッダとして“To:”、“Subject:”および“From:”の各行が付加されている。

“To:”行にはテキストを配布する宛て先、例えば電子メールの送り先のアドレスを付加する。図20では、ユーザ1およびユーザ3にテキストを配布するため、“ユーザ1”および“ユーザ3”と記述されている。

“Subject:”行にはユーザが識別しやすい情報を付加する。本図では、配布するテキストの最初の数文字を抜き出して記述しているが、ここには何を付加してもよい。

“From:”行にはテキストの送り元、例えば電子メールの送り元のアドレスを付加する。本図では、テキス

トを配布するシステムの名称である“文書検索配送システム”と記述されている。以上が本実施例におけるテキスト検索制御プログラム112におけるテキスト検索の詳細な手順である。

【0054】テキスト検索制御プログラム112の処理の結果、成立している検索条件式が一つでも存在した場合には、図7に示したように、テキスト検索制御プログラム112の終了後、電子メールプログラム119がシステム制御プログラム105によって起動される。本プログラムでは、テキスト成形プログラム118によって付加されたヘッダを参照して、テキストを電子メールによって配布する。例えば、図20の例の場合、ヘッダの“To:”行に相当する部分を参照し、そこに記述されている宛て先にヘッダが付加されたテキストを送付する。本図では、“To:”行に“ユーザ1”および“ユーザ3”と記述されているため、ユーザ1とユーザ3にテキストを配布し、処理を終了する。

【0055】以上説明したように、本発明によれば、複数ユーザの検索条件式の登録時に、それらの検索条件式に含まれる検索タームを指定したユーザの識別情報とそのユーザが指定した検索条件式に含まれる検索タームの数を記憶しておき、テキストの検索時に、テキスト中に一致する部分文字列が出現した検索タームのユーザ毎の数と記憶しておいたユーザ毎の検索ターム数とを比較することによって全ての検索条件式が成立しているかどうかを判別することができるため、ただ一度のテキスト走査で全てのユーザの検索条件式の成否を判定でき、全ユーザの検索条件式に関する検索処理を一度に行うことが可能となる。その結果、複数の情報源から入手したテキストを、ユーザが予め登録しておいた検索条件式に基づき、テキストのただ一度の走査で複数ユーザの検索条件式が成立しているかどうかを判別し、条件が成立しているユーザに対して即座にそのテキストを配布することができる即時性の高い文書検索配送システムを実現することが可能となる。また、この文書検索配送システムは即時性が高いので、ユーザがシステムに検索条件式を通知してから検索されたテキストが配送されるまでの時間が短く、この時間を監視することにより本発明が適用されているか否かを判定することが可能である。

【0056】《第二実施例》次に、本発明の第二の実施例について説明する。本実施例で示す文書検索配送システムでは、ユーザ毎に配布条件を管理することにより、ある程度まとめてテキストを配布したり、決まった時間に配布するなど、ユーザの希望に応じてテキストを配布することが可能となる。また、商業的なシステムとして用いる場合には、ユーザの契約条件に応じて時間遅延を設けてテキストを配布することも可能となる。

【0057】本実施例は第一の実施例(図1)と基本的に同様の構成をとるが、その中の主メモリ104内の構成が異なる。この主メモリ104内の構成は図21に示

すようなものとなる。図21に示すように、主メモリ104aに配布管理テーブル2108を確保し、システム制御プログラム105aの制御下に配布条件登録制御プログラム2100およびテキスト配布制御プログラム2104を新たに設ける。また、配布条件登録制御プログラム2100の制御下に配布条件解析プログラム2101および配布条件登録プログラム2102を、テキスト検索制御プログラム112aの制御下に配布情報格納プログラム2103を、テキスト配布制御プログラム2104の制御下にテキスト配布プログラム2105を設ける。このテキスト配布プログラム2105は配布条件チェックプログラム2106、電子メールプログラム119および配布情報修正プログラム2107で構成される。電子メールプログラム119にはワークステーションなどで一般的に用いられている既存のメールプログラムを用いる。以上のプログラムはハードディスク装置、フロッピーディスクなどのコンピュータで読み書きできる記憶媒体に格納することもできる。

【0058】システム制御プログラム105aは文書検索配送システムの管理者によるキーボード101からの指示を受けて起動される。配布条件登録制御プログラム2100、検索条件式登録制御プログラム106、テキスト検索制御プログラム112aおよびテキスト配布制御プログラム2104はユーザ126からの配布条件や検索条件式の登録指示、キーボード101からの配布条件の登録指示およびニュース配信元125からのテキストの配信により、システム制御プログラム105aによって起動され、それぞれ配布条件解析プログラム2101および配布条件登録プログラム2102の制御、検索条件式解析プログラム107、検索ターム数カウントテーブル作成プログラム108および検索用オートマトン作成プログラム109の制御、テキスト取得プログラム113、テキスト検索プログラム114、テキスト成形プログラム118および配布情報格納プログラム2103の制御、テキスト配布プログラム2105の制御を行う。以下、本実施例における文書検索配送システムの処理内容について説明する。

【0059】まず、システム制御プログラム105aの処理内容について図22のPAD図を用いて説明する。システム制御プログラム105aでは、まずステップ2200で、キーボード101から終了コマンドが入力されるまで、以下のステップを繰り返す。この繰り返し処理では、まずステップ2201でユーザ126からの電子メールあるいはキーボード101の入力によって配布条件が送られてきているかどうかを調べる。ここで、配布条件が送られてきている場合には、ステップ2202で配布条件登録制御プログラム2100を起動して、配布条件の登録を行う。次に、ステップ2203でユーザ126から電子メールによって検索条件式が送られてきているかどうかを調べる。ここで、検索条件式が送られ

てきている場合には、ステップ2204で検索条件式登録制御プログラム106を起動して、検索条件式の登録を行う。次に、ステップ2205でニュース配信元125から電子メールによってテキストが送られてきているかどうかを調べる。ここで、テキストが送られてきている場合には、ステップ2206でテキスト検索制御プログラム112aを起動して、テキストの検索を行う。最後に、ステップ2207でテキスト配布制御プログラム2104を起動し、配布条件を判定してその条件を満たしているユーザに対してのみテキストを配布する。以上がシステム制御プログラム105aの処理内容である。

【0060】以下、第一の実施例にはない配布条件登録制御プログラム2100とテキスト配布制御プログラム2104および第一の実施例と処理が異なるテキスト検索制御プログラム112aの処理内容について説明する。まず、配布条件登録制御プログラム2100による配布条件登録の処理内容について図23のPAD図を用いて説明する。配布条件登録制御プログラム2100はシステム制御プログラム105aによって起動される。本プログラムは、まずステップ2300で配布条件解析プログラム2101を起動し、ユーザ126からの電子メールあるいはキーボード101の入力によって送られてきた配布条件を解析する。この配布条件の解析処理では、配布条件から以下の情報を抽出する。

- (A) 配布条件を設定するユーザの識別子
- (B) 配布条件の形式
- (C) 配布条件の設定値

上記(B)の配布条件の形式としては、「配布時間」、「配布件数」、「遅延時間」などの配布条件の種別を抽出する。(C)の配布条件の設定値として抽出する値は、例えば(B)が「配布時間」ならその時間、「配布件数」なら配布する件数、「遅延時間」なら検索してから実際に配布するまでの経過時間である。例えば、ユーザ番号1：配布時間(18:00)という「ユーザ番号1に対して18:00に配布する」ことを意味する配布条件が送られてきた場合には「ユーザ番号1」、「配布時間」および「18:00」を抽出する。

ユーザ番号2：配布件数(5)

という「ユーザ番号2に対して5件たまったら配布する」ことを意味する配布条件が送られてきた場合には「ユーザ番号2」、「配布件数」および「5」を抽出する。

ユーザ番号3：遅延時間(01:30)

という「ユーザ番号3に対して1時間30分遅れて配布する」ことを意味する配布条件が送られてきた場合には「ユーザ番号3」、「遅延時間」および「01:30」を抽出する。最後に、ステップ2301で配布条件登録プログラム2102を起動し、配布条件解析プログラム2101によって解析された結果を配布管理テーブル2



108に格納する。図24に配布管理テーブル2108の例を示す。配布管理テーブル2108には配布条件解析プログラム2101によって抽出された配布条件の形式と設定値がユーザ番号に対応する形で格納され、配布条件チェック用領域および配布テキスト番号格納用領域が確保されている。

【0061】次に、テキスト検索制御プログラム112aによるテキスト検索の処理内容について図25のPAD図を用いて説明する。テキスト検索制御プログラム112aはシステム制御プログラム105aによって起動される。図25に示す本プログラムの処理内容のうち、ステップ2500～2503におけるテキスト取得プログラム113、テキスト検索プログラム114およびテキスト成形プログラム118の処理内容は第一の実施例で述べた通りである。テキスト検索制御プログラム112aはステップ2504で、配布情報格納プログラム2103を起動し、配布管理テーブル2108の配布テキスト番号格納用領域に検索条件が成立した配布すべきテキストの番号を追加格納する。次に、ステップ2505で配布管理テーブル2108の配布テキスト番号格納用領域に格納してあるテキスト番号の数、あるいは現在の時間を本テーブルの配布条件チェック用領域に格納する。このとき、配布条件の形式が「配布件数」の場合には格納してあるテキスト番号の数を、「遅延時間」の場合には現在の時間を格納する。「配布時間」の場合には何も格納する必要はない。その後、ステップ2506でテキスト成形プログラム118によって成形されたテキストをワークエリア123に格納する。

【0062】最後に、テキスト配布制御プログラム2104によるテキスト配布の処理内容について図26のPAD図を用いて説明する。テキスト配布制御プログラム2104はシステム制御プログラム105aによって起動される。本プログラムは、ステップ2600でテキスト配布プログラム2105を起動し、ユーザ毎に配布条件を判定し、その条件を満たしているユーザに対してテキストを配布する。

【0063】テキスト配布プログラム2105の詳細な処理内容を図27に示す。テキスト配布プログラム2105は、まずステップ2700で、配布管理テーブル2108に配布条件が格納されている全てのユーザ番号に対して以下のステップを繰り返す。この繰り返し処理では、まずステップ2701で配布条件チェックプログラム2106を起動し、配布管理テーブル2108を用いて配布条件を満たしているか否かを判定する。ここで、以下の条件を満たしていれば配布条件を満たしているとみなす。

条件(1)：配布条件の形式が「配布時間」で、配布条件の設定値と現在の時間が一致するか、あるいは配布条件の設定値より現在の時間の方が超過している。

条件(2)：配布条件の形式が「配布件数」で、配布条

件の設定値と配布条件チェック用領域に格納されている件数が一致する。

条件(3)：配布条件の形式が「遅延時間」で、配布条件の設定値と配布条件チェック用領域に格納されている時間から現在の時間までの経過時間が一致するか、あるいは配布条件の設定値より経過時間の方が超過している。上記の条件を満たしている場合には、ステップ2702で配布条件を満たしていると判断し、ステップ2703で電子メールプログラム119を起動して配布管理テーブル2108の配布テキスト番号格納用領域に格納されている番号のテキストをそのユーザ番号に配布する。最後に、ステップ2704で配布情報修正プログラム2107を起動し、テキストを配布したユーザ番号に対応する配布管理テーブル2108の配布条件チェック用領域と配布テキスト番号格納用領域をリセットする。これは、配布条件チェック用領域をNULLクリアし、配布テキスト番号格納用領域からテキスト番号を削除することで実現する。以上が本発明の文書検索システムの実施例である。

【0064】以下、図23に示した本実施例における配布条件登録制御プログラム2100の処理手順について図28を用いて具体的に説明する。まず、図23の配布条件登録制御プログラム2100のステップ2300における配布条件解析プログラム2101の処理について説明する。配布条件解析プログラム2101は配布条件登録制御プログラム2100によって起動される。本プログラムは、ユーザ126から電子メールで送られてきた配布条件あるいはキーボード101から入力された配布条件を解析する。例として「ユーザ番号1：配布時間(18:00)」、「ユーザ番号2：配布件数(5)」および「ユーザ番号3：遅延時間(01:30)」という配布条件を解析した結果を図28に示す。例えば「ユーザ番号1：配布時間(18:00)」という配布条件の場合、解析結果として、配布条件を設定するユーザ番号「1」、配布条件の形式「配布時間」、配布条件の設定値「18:00」が得られる。

【0065】次に、図23の配布条件登録制御プログラム2100のステップ2301における配布条件登録プログラム2102の処理について説明する。配布条件登録プログラム2102は配布条件登録制御プログラム2100によって、配布条件解析プログラム2101の次に起動される。本プログラムは、配布条件解析プログラム2101による解析結果に基づき配布管理テーブル2108を作成する。作成された配布管理テーブル2108の例を図28に示す。本テーブルには、配布条件解析プログラム2101による解析結果に基づき、各ユーザ番号に対応して配布条件の形式および設定値を格納する。また、配布条件チェック用領域および配布テキスト番号格納用領域を確保する。以上が本実施例における配布条件登録制御プログラム2100における検索条件式

登録処理の詳細な手順である。

【0066】以下、図25に示した本実施例におけるテキスト検索制御プログラム112aの処理手順について具体的に説明する。図25に示す本プログラムの処理内容のうち、ステップ2500～2503におけるテキスト取得プログラム113、テキスト検索プログラム114およびテキスト成形プログラム118の処理内容は第一の実施例で詳しく述べた通りである。以下は、ステップ2504～2506における配布情報格納プログラム2103の詳細な処理内容である。配布情報格納プログラム2103は、テキスト検索制御プログラム112aによってテキスト成形プログラム118の次に起動される。本プログラムは、まず、ステップ2504でテキストの番号を、検索条件式が成立したユーザ番号に対応する配布管理テーブル2108の配布テキスト番号格納用領域に格納する。図29に本プログラムの処理内容の例を示す。本図は、59番のテキストに対してユーザ番号1およびユーザ番号2の検索条件式が成立した場合の例である。そのため、配布管理テーブル2108の配布テキスト番号格納用領域のユーザ番号1およびユーザ番号2に対応する場所にテキスト番号“59”が格納される。次に、配布情報格納プログラム2103は、ステップ2504で配布管理テーブル2108の配布テキスト番号格納用領域に格納してあるテキスト番号の数、あるいは現在の時間を配布管理テーブル2108の配布条件チェック用領域に格納する。このとき、配布条件の形式が「配布件数」の場合には格納してあるテキスト番号の数を、「遅延時間」の場合には現在の時間を格納する。「配布時間」の場合には何も格納する必要はない。図29の例の場合、ユーザ番号2の配布条件の形式は「配布件数」であるので、配布条件チェック用領域の値を1増やして“5”にするが、ユーザ番号1の配布条件の形式は「配布時間」であるため何もしない。最後に、配布情報格納プログラム2103はステップ2506で、テキスト成形プログラム118によって成形されたテキストをワークエリア123に格納して終了する。

【0067】最後に、図26のテキスト配布制御プログラム2104のステップ2600におけるテキスト配布プログラム2105の処理について説明する。テキスト配布プログラム2105の詳細な処理内容は図27に示した通りである。まず、ステップ2700で、配布管理テーブル2108に配布条件が格納されている全てのユーザに対して以下の処理を繰り返す。この繰り返し処理では、まずステップ2701で配布条件チェックプログラム2105を起動し、配布条件を判定する。ここで、以下の条件を満たしていれば配布条件を満たしているとみなす。

条件(1)：配布条件の形式が「配布時間」で、配布条件の設定値と現在の時間が一致するか、あるいは配布条件の設定値より現在の時間の方が超過している。

条件(2)：配布条件の形式が「配布件数」で、配布条件の設定値と配布条件チェック用領域に格納されている件数が一致する。

条件(3)：配布条件の形式が「遅延時間」で、配布条件の設定値と配布条件チェック用領域に格納されている時間から現在の時間までの経過時間が一致するか、あるいは配布条件の設定値より経過時間の方が超過している。

【0068】ユーザ番号1の配布条件の形式は「配布時間」である。しかし、現在の時間“14:00”は配布条件の設定値“18:00”を超過していないため、配布条件が満たされていない。そのため、次の繰り返しに移る。

【0069】ユーザ番号2の配布条件の形式は「配布件数」であり、配布条件チェック用領域に格納されている件数と配布条件の設定値が“5”と一致しているので、ステップ2703で電子メールプログラム119を起動し、配布管理テーブル2108の配布テキスト番号格納用領域に格納されている番号のテキストを配布する。本図においてユーザ番号2に対応する配布テキスト番号格納用領域には、テキスト番号19、24、33、42、59が格納されているので、ワークエリア123に格納されているテキストで同じ番号のものをユーザ番号2に配布する。次に、ステップ2704で配布情報修正プログラム2107を起動し、テキストを配布したユーザ番号、すなわちユーザ番号2に対応する配布管理テーブル2108の配布チェック用領域および配布テキスト番号格納用領域をリセットし、次の繰り返し処理を行う。

【0070】ユーザ番号3の配布条件の形式は「遅延時間」であり、配布条件チェック用領域に格納されている時間から現在の時間までの経過時間と配布条件の設定値に格納されている時間が“01:30”と一致しているので、ステップ2703で電子メールプログラム119を起動し、配布管理テーブル2108の配布テキスト番号格納用領域に格納されている番号のテキストを配布する。本図においてユーザ番号3に対応する配布テキスト番号格納用領域には、テキスト番号53が格納されているので、ワークエリア123に格納されているテキストで同じ番号のものをユーザ番号3に配布する。次に、ステップ2704で配布情報修正プログラム2107を起動し、テキストを配布したユーザ番号、すなわちユーザ番号2に対応する配布管理テーブル2108の配布チェック用領域および配布テキスト番号格納用領域をリセットする。全ての繰り返し処理が終了したときの配布管理テーブル2108を図30に示す。ユーザ番号2およびユーザ番号3にテキスト配布処理を行ったので、それらに対応する配布条件チェック用領域および配布テキスト番号格納用領域がリセットされている。

【0071】以上説明したように、本発明によれば、ユーザ毎に配布条件を設定し、その配布条件にしたがって

検索条件式が成立したテキストを配布することにより、ある程度まとめてテキストを配布したり、決まった時間に配布するなど、ユーザの希望に応じてテキストを配布することが可能となる。また、商業的なシステムとして用いる場合には、ユーザの契約条件に応じて時間遅延を設けてテキストを配布することも可能となる。その結果、複数の情報源から入手したテキストを、ユーザが予め登録しておいた検索条件式に基づき、テキストのただ一度の走査で複数ユーザの検索条件式が成立しているかどうかを判別し、ユーザの個々の希望配布条件に応じてそのテキストを配布することができる柔軟性の高い文書検索配送システムを実現することが可能となる。

【0072】《第三実施例》次に、本発明の第三の実施例について説明する。本実施例で示す文書検索配送システムでは、各ユーザが指定した検索条件式に含まれる検索タームをユーザ毎に管理し、ユーザから検索条件式の削除が指示された際に、管理してある検索タームで有限オートマトンを探索し、ユーザリストのポイントを張り替えることによって、容易にユーザリストから以前の情報を削除することができる。この実施例によれば、ユーザによる検索条件式の変更指示がある場合にも容易に変更を行うことが可能となる。本実施例は第一の実施例

(図1)と基本的に同様の構成をとるが、その中の主メモリ104内の構成が異なる。この主メモリ104内の構成は図31に示すようなものとなる。図31に示すように、主メモリ104bに検索ターム管理テーブル3106を確保し、システム制御プログラム105bの制御下に検索条件式削除制御プログラム3100を新たに設ける。また、検索条件式削除制御プログラム3100の制御下にユーザリスト修正プログラム3101および検索ターム数カウントテーブル修正プログラム3104を、検索条件式登録制御プログラム106bの制御下に検索ターム管理テーブル作成プログラム3105を設ける。ユーザリスト修正プログラム3101は有限オートマトン探索プログラム3102およびユーザリスト部分削除プログラム3103で構成される。以上のプログラムはハードディスク装置、フロッピーディスクなどのコンピュータで読み書きできる記憶媒体に格納することもできる。

【0073】システム制御プログラム105bは文書検索配送システムの管理者によるキーボード101からの指示で起動される。検索条件式削除制御プログラム3100、検索条件式登録制御プログラム106bおよびテキスト検索制御プログラム112はユーザ126からの検索条件式の登録や削除指示およびニュース配信元125からのテキストの配信により、システム制御プログラム105bによって起動され、それぞれユーザリスト修正プログラム3101および検索ターム数カウントテーブル修正プログラム3104の制御、検索条件式解析プログラム107、検索ターム数カウントテーブル作成プ

ログラム108、検索用オートマトン作成プログラム109および検索ターム管理テーブル作成プログラム3105の制御、テキスト取得プログラム113、テキスト検索プログラム114およびテキスト成形プログラム118の制御を行う。

【0074】以下、本実施例における文書検索配送システムの処理内容について説明する。まず、システム制御プログラム105bの処理内容について図32のPAD図を用いて説明する。システム制御プログラム105bでは、まずステップ3200で、キーボード101から終了コマンドが入力されるまで、以下のステップを繰り返す。この繰り返し処理では、まずステップ3201でユーザ126から電子メールによって検索条件式の削除指示が送られてきているかどうかを調べる。ここで、検索条件式の削除指示が送られてきている場合には、ステップ3202で検索条件式削除制御プログラム3100を起動して、検索条件式の削除を行う。次に、ステップ3203でユーザ126から電子メールによって検索条件式が送られてきているかどうかを調べる。ここで、検索条件式が送られてきている場合には、ステップ3204で検索条件式登録制御プログラム106bを起動して、検索条件式の登録を行う。次に、ステップ3205でニュース配信元125から電子メールによってテキストが送られてきているかどうかを調べる。ここで、テキストが送られてきている場合には、ステップ3206でテキスト検索制御プログラム112を起動して、テキストの検索を行う。次に、ステップ3207でテキスト検索制御プログラム112におけるテキスト検索の結果を調べ、成立している検索条件式が一つでも存在すると判断された場合には、ステップ3208で電子メールプログラム119を起動し、成立した検索条件式を指定したユーザに対し、該当テキストを電子メールを用いて配布する。以上がシステム制御プログラム105bの処理内容である。

【0075】以下、第一の実施例にはない検索条件式削除制御プログラム3100および第一の実施例と処理が異なる検索条件式登録制御プログラム106bの処理内容について説明する。まず、検索条件式削除制御プログラム3100による検索条件式の削除処理の内容について図33のPAD図を用いて説明する。検索条件式削除制御プログラム3100はシステム制御プログラム105bによって起動される。本プログラムは、まずステップ3300でユーザリスト修正プログラム3101を起動し、検索条件式の削除が指定されたユーザ番号に対応するユーザリスト122を削除する。そして、ステップ3301で検索ターム数カウントテーブル修正プログラム3104を起動し、検索条件式の削除が指定されたユーザ番号に対応する検索ターム数カウントテーブル120を削除する。

【0076】ユーザリスト修正プログラム3101の詳細

細な処理内容を図34に示す。本プログラムは、まずステップ3400で検索ターム管理テーブル3106に格納されている検索タームの数分、以下のステップを繰り返す。検索ターム管理テーブル3106については後で説明する。繰り返し処理では、まずステップ3401で有限オートマトン探索プログラム3102を起動し、検索タームで有限オートマトン121を探索することによってユーザリスト122へのポインタを得る。最後に、ステップ3402でユーザリスト部分削除プログラム3103を起動し、ユーザリスト122のポインタを張り替えることによって、検索条件式の削除を行うユーザ番号に対応するリストを削除する。以上が、検索条件式削除制御プログラム3100による検索条件式削除の処理内容である。

【0077】次に、検索条件式登録制御プログラム106bによる検索条件式の登録処理の内容について図35のPAD図を用いて説明する。検索条件式登録制御プログラム106bはシステム制御プログラム105bによって起動される。図35に示す本プログラムの処理内容のうち、ステップ3500～3502における検索条件式解析プログラム107、検索ターム数カウントテーブル作成プログラム108および検索用オートマトン作成プログラム109の処理内容については第一の実施例で述べた通りである。検索条件式登録制御プログラム106bはステップ3503で、検索ターム管理テーブル作成プログラム3105を起動し、検索条件式中に含まれる検索タームを検索ターム管理テーブル3106に格納する。ここで、作成される検索ターム管理テーブル3106の例を図36に示す。本図に示した検索ターム管理テーブル3106は、「“文書” and (“検索” or “サーチ”）」というユーザ1の検索条件式、「“文字” and “(“認識” or “学習”）」というユーザ2の検索条件式および「“(“検索” and “学習”）」というユーザ3の検索条件式から抽出した検索タームから作成されたものである。以上が本発明の文書検索方法の実施例である。

【0078】以下、図33に示した本実施例における検索条件式削除制御プログラム3100の処理手順について図37を用いて具体的に説明する。まず、図33の検索条件式削除制御プログラム3100のステップ3300におけるユーザリスト修正プログラム3101の処理について説明する。ユーザリスト修正プログラム3101は検索条件式削除制御プログラム3100によって起動される。本プログラムの詳細な処理内容は図34に示した通りである。本プログラムはステップ3400で、検索ターム管理テーブル3106に格納されている検索タームの数分、ステップ3401およびステップ3402を繰り返す。ステップ3401では、有限オートマトン探索プログラム3102を起動し、検索タームで有限オートマトン121を探索してユーザリスト122への

ポインタを得る。ステップ3402では、ユーザリスト部分削除プログラム3103を起動し、検索条件式の削除が指定されたユーザ番号に対応するユーザリスト122を削除する。

【0079】ユーザリスト修正プログラム3101の処理例を図37に示す。本図は「文字」、「認識」および「学習」を検索条件式中含むユーザ番号2の検索条件式を削除する場合の例である。検索ターム管理テーブル3106には「文字」、「認識」および「学習」が格納されており、これらの検索タームに対して繰り返し処理が行われる。繰り返し処理の中でステップ3401では「文字」、「認識」および「学習」で有限オートマトン121を探索し、それぞれユーザリスト122へのポインタを得る。ステップ3402では、それらのユーザリスト122を辿り、ユーザ番号2に関するユーザリスト122を削除する。本図では、ユーザリスト122のうちユーザ番号“2”の部分が削除される。

【0080】次に、図33の検索条件式削除制御プログラム3100のステップ3301における検索ターム数カウントテーブル修正プログラム3104の処理について説明する。検索ターム数カウントテーブル修正プログラム3104は、検索条件式削除制御プログラム3100によってユーザリスト修正プログラム3101の次に起動される。本プログラムでは、検索条件式の削除が指定されたユーザ番号に対応する検索ターム数カウントテーブル120を削除する。図38に本プログラムの処理例を示す。本図に示す通り、ユーザ番号2に対応する検索ターム数カウントテーブルが削除される。以上が本実施例における検索条件式削除制御プログラム3100における検索条件式削除の詳細な手順である。

【0081】以下、図35に示した本実施例における検索条件式登録制御プログラム106bの処理手順について具体的に説明する。図35に示す本プログラムの処理内容のうち、ステップ3500～3502における検索条件式解析プログラム107、検索ターム数カウントテーブル作成プログラム108および検索用オートマトン作成プログラム109の処理内容については第一の実施例で述べた通りである。以下、ステップ3503における検索ターム管理テーブル作成プログラム3105の詳細な処理内容を説明する。検索ターム管理テーブル作成プログラム3105は、検索条件式登録制御プログラム106bによって検索用オートマトン作成プログラム109の次に起動される。本プログラムでは、検索条件式解析プログラム107による解析結果として得られた検索条件式中に含まれる検索タームを検索ターム管理テーブル3106に格納する。

【0082】本プログラムの処理例を図39に示す。本図は、ユーザ番号2のユーザが「“構造” and “認識”」、すなわち「“構造”と“認識”を同時に含む文書」という検索条件式を指定した場合の例である。この

検索条件式を検索条件式解析プログラム 107 によって解析すると“構造”および“認識”という 2 つの検索タームが得られる。これらの検索タームを検索ターム管理テーブル 3106 にユーザ番号に対応した形で格納する。本図では、ユーザ番号 2 に対応する場所に“構造”および“認識”が格納されている。以上が本実施例における検索条件式登録制御プログラム 106 b における検索条件式登録の詳細な手順である。

【0083】以上説明したように、本発明によれば、各ユーザが指定した検索条件式に含まれる検索タームをユーザ毎に管理し、ユーザから検索条件式の削除が指示された際に、管理してある検索タームで有限オートマトンを探索し、ユーザリストのポインタを張り替えることによって、容易にユーザリストから以前の情報を削除することが可能となる。その結果、複数の情報源から入手したテキストを、ユーザが予め登録しておいた検索条件式に基づき、テキストのただ一度の走査で複数ユーザの検索条件式が成立しているかどうかを判別し、条件が成立しているユーザに対して即座にそのテキストを配布する文書検索配送システムにおいて、ユーザの要求に応じて検索条件式を随時変更できる文書検索配送システムを実現することが可能となる。本実施例では、検索条件式の削除指示と登録指示が別々に送られてくる場合について説明したが、検索条件式の削除指示と登録指示が同時に送られてくる、すなわち更新指示が送られてくる場合にも削除処理と登録処理を連続して行うことにより対応できることは明らかであろう。また、本実施例は第一の実施例に対して検索条件式の削除処理を追加したものであるが、第二の実施例に対して適用することも可能なことは明らかであろう。

#### 【0084】

【発明の効果】本発明によれば、ユーザ数が、すなわち検索条件式の数的大量となる場合でも、ただ一度のテキスト走査で全ての検索条件式が成立しているかどうかを判別することができるため、高速なテキスト検索を実現することが可能となる。また、この結果、ユーザ数が増えた場合でも、ユーザ数に依らない高速なリアルタイムのテキスト検索配布が行える文書検索配送システムを提供することが可能となる。

#### 【図面の簡単な説明】

【図 1】本発明の第一の実施例の構成を示す図である。

【図 2】従来技術 1 における有限オートマトンの構成を示す図である。

【図 3】従来技術 1 におけるフェイル先状態番号テーブルの構成を示す図である。

【図 4】従来技術 1 における出力テーブルの構成を示す図である。

【図 5】従来技術 2 における日本語テキスト用有限オートマトンの例を示す図である。

【図 6】本発明の概略の処理内容を示す図である。

【図 7】システム制御プログラム 105 の処理内容を示す PAD 図である。

【図 8】検索条件式登録制御プログラム 106 の処理内容を示す PAD 図である。

【図 9】検索条件式の解析方法を説明するための図である。

【図 10】検索ターム数カウントテーブル 120 の作成方法を説明するための図である。

【図 11】テキスト検索制御プログラム 112 の処理内容を示す PAD 図である。

【図 12】検索用オートマトン作成プログラム 109 の処理内容を示す PAD 図である。

【図 13】有限オートマトン 121 およびユーザリスト 122 の作成方法を説明するための図である。

【図 14】テキスト検索プログラム 114 の処理内容を示す PAD 図である。

【図 15】ユーザリスト 122 の作成処理を説明するための図である。

【図 16】テキスト走査処理を説明するための図である。

【図 17】テキスト中に一致する部分文字列が出現した検索タームのカウント処理を説明するための図である。

【図 18】テキスト中に一致する部分文字列が出現しなかった検索タームのカウント処理を説明するための図である。

【図 19】検索条件式の成立チェック処理を説明するための図である。

【図 20】テキスト成形処理を説明するための図である。

【図 21】本発明の第二の実施例の構成を示す図である。

【図 22】システム制御プログラム 105 a の処理内容を示す PAD 図である。

【図 23】配布条件登録制御プログラム 2100 の処理内容を示す PAD 図である。

【図 24】配布管理テーブル 2108 の構成を示す図である。

【図 25】テキスト検索制御プログラム 112 a の処理内容を示す PAD 図である。

【図 26】テキスト配布制御プログラム 2104 の処理内容を示す PAD 図である。

【図 27】テキスト配布プログラム 2105 の処理内容を示す PAD 図である。

【図 28】配布条件登録処理を説明するための図である。

【図 29】配布情報格納処理を説明するための図である。

【図 30】配布条件チェック処理および配布情報修正処理を説明するための図である。

【図 31】本発明の第三の実施例の構成を示す図であ

る。

【図32】システム制御プログラム105bの処理内容を示すPAD図である。

【図33】検索条件式削除制御プログラム3100の処理内容を示すPAD図である。

【図34】ユーザリスト修正プログラム3101の処理内容を示すPAD図である。

【図35】検索条件式登録制御プログラム106bの処理内容を示すPAD図である。

【図36】検索ターム管理テーブル3106の構成を示す図である。

【図37】ユーザリスト修正処理を説明するための図である。

【図38】検索ターム数カウントテーブル修正処理を説明するための図である。

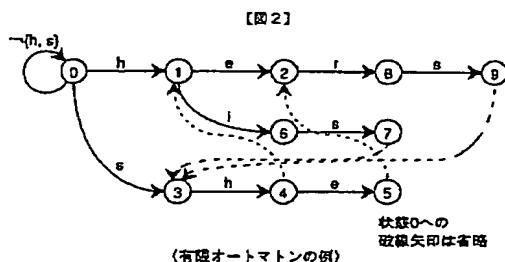
【図39】検索ターム管理テーブル作成処理を説明するための図である。

【符号の説明】

- 100 ディスプレイ
- 101 キーボード
- 102 CPU
- 103 バス
- 104、104a、104b 主メモリ
- 105、105a、105b システム制御プログラム
- 106、106b 検索条件式登録制御プログラム
- 107 検索条件式解析プログラム
- 108 検索ターム数カウントテーブル作成プログラム
- 109 検索用オートマトン作成プログラム
- 110 有限オートマトン作成プログラム
- 111 ユーザリスト作成プログラム
- 112、112a テキスト検索制御プログラム

- 113 テキスト取得プログラム
- 114 テキスト検索プログラム
- 115 テキスト走査プログラム
- 116 検索ターム数カウントプログラム
- 117 検索条件式チェックプログラム
- 118 テキスト成形プログラム
- 119 電子メールプログラム
- 120 検索ターム数カウントテーブル
- 121 有限オートマトン
- 122 ユーザリスト
- 123 ワークエリア
- 124 LAN
- 125 ニュース配信元
- 126 文書検索配送システムのユーザ
- 2100 配布条件登録制御プログラム
- 2101 配布条件解析プログラム
- 2102 配布条件登録プログラム
- 2103 配布情報格納プログラム
- 2104 テキスト配布制御プログラム
- 2105 テキスト配布プログラム
- 2106 配布条件チェックプログラム
- 2107 配布情報修正プログラム
- 2108 配布管理テーブル
- 3100 検索条件式削除制御プログラム
- 3101 ユーザリスト修正プログラム
- 3102 有限オートマトン探索プログラム
- 3103 ユーザリスト部分削除プログラム
- 3104 検索ターム数カウントテーブル修正プログラム
- 3105 検索ターム管理テーブル作成プログラム
- 3106 検索ターム管理テーブル

【図2】



【図3】

【図3】

状態番号	1	2	3	4	5	6	7	8	9
フェイルによる遷移先	0	0	0	1	2	0	3	0	3

〈フェイル先状態番号テーブルの例〉

【図4】

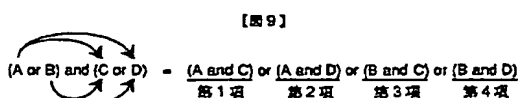
【図4】

2	he
5	she, he
7	his
8	here

↑ 状態番号    ↑ 出力

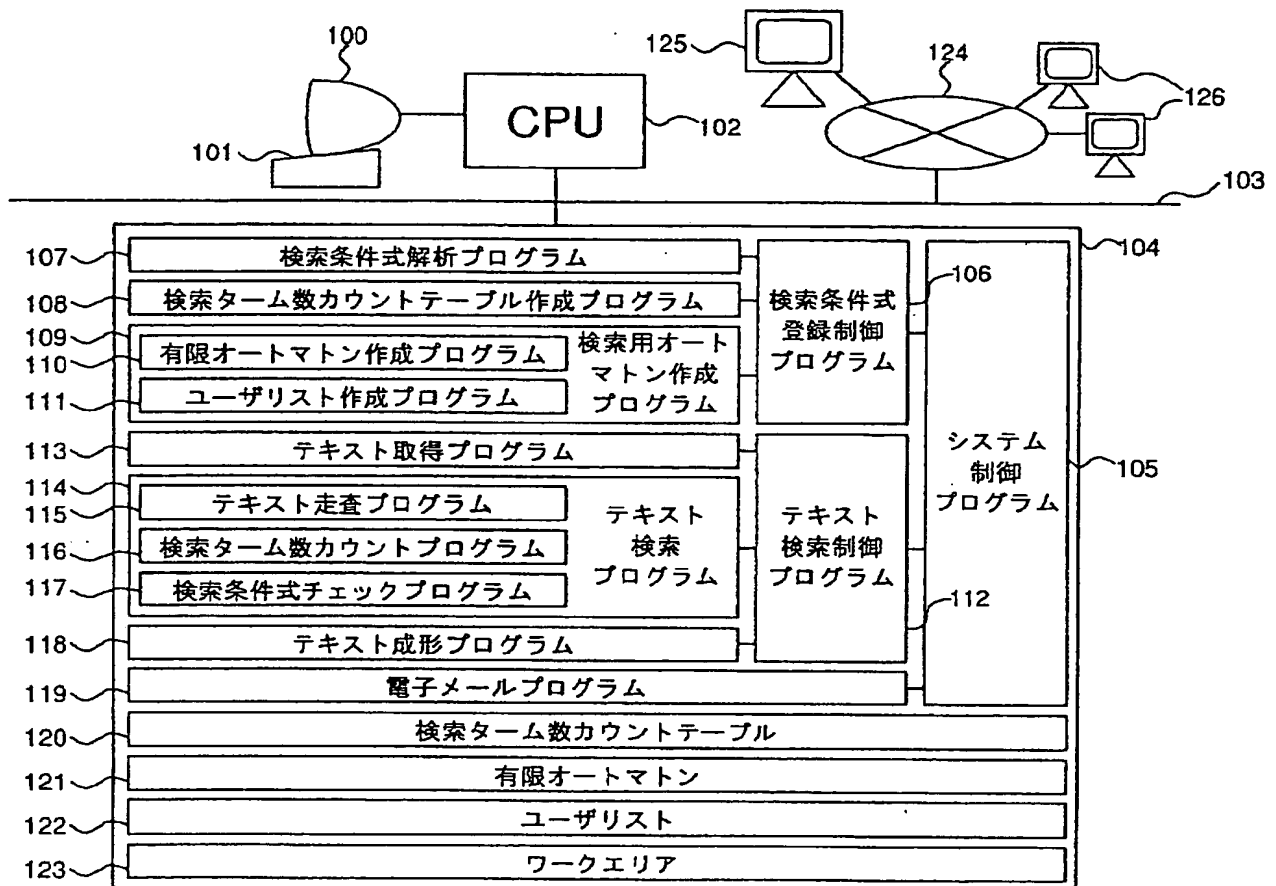
〈出力テーブルの例〉

【図9】



【図 1】

【図 1】

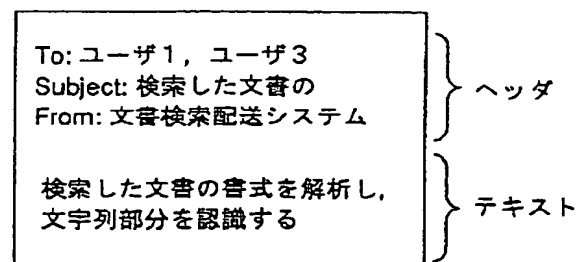
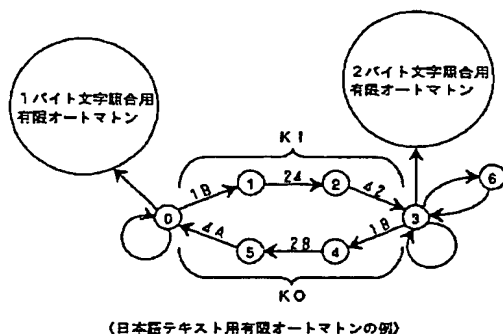


【図 5】

【図 20】

【図 5】

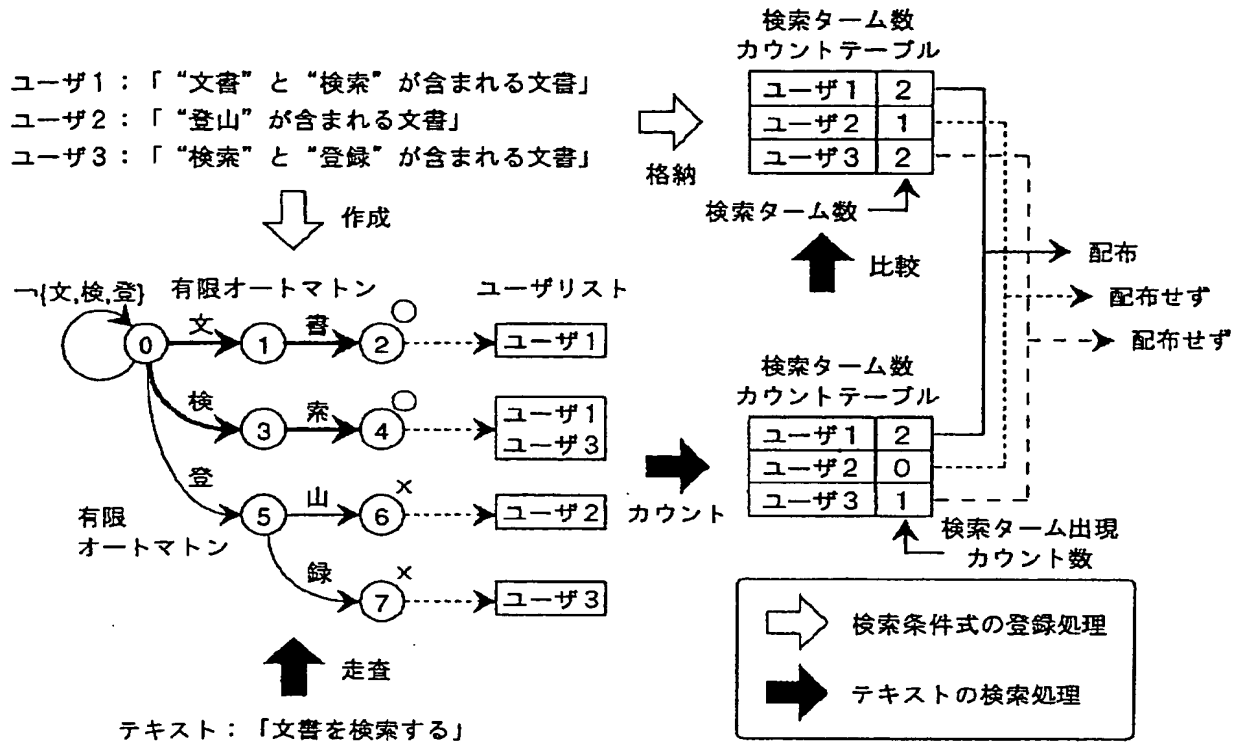
【図 20】



〈成形テキストの例〉

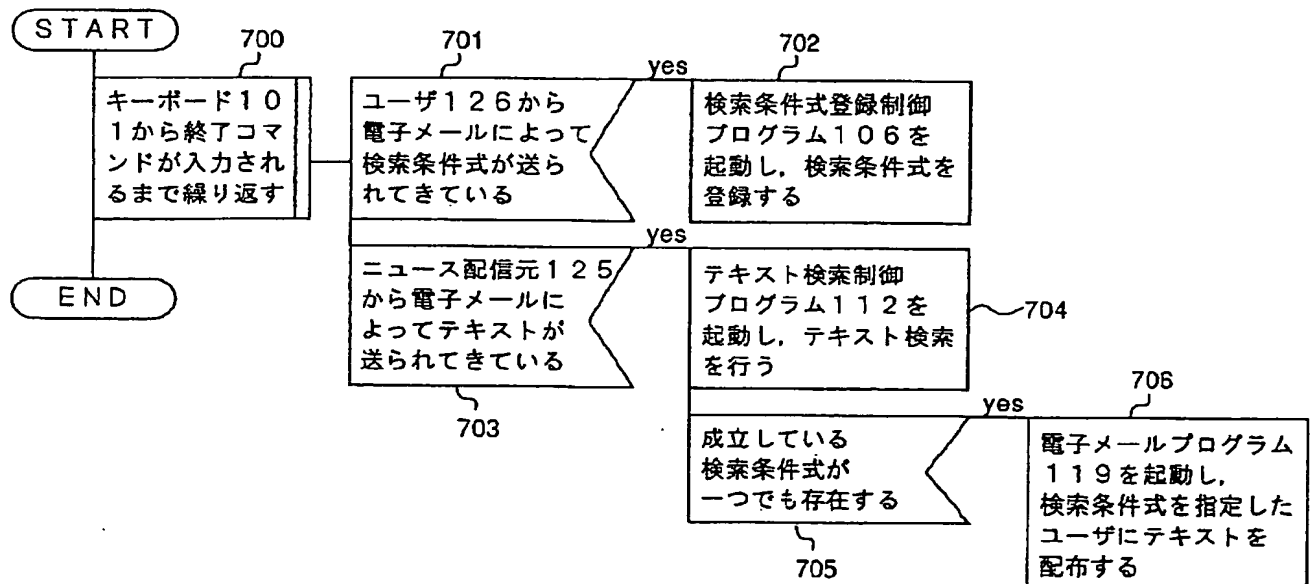
【図6】

【図6】



【図7】

【図7】

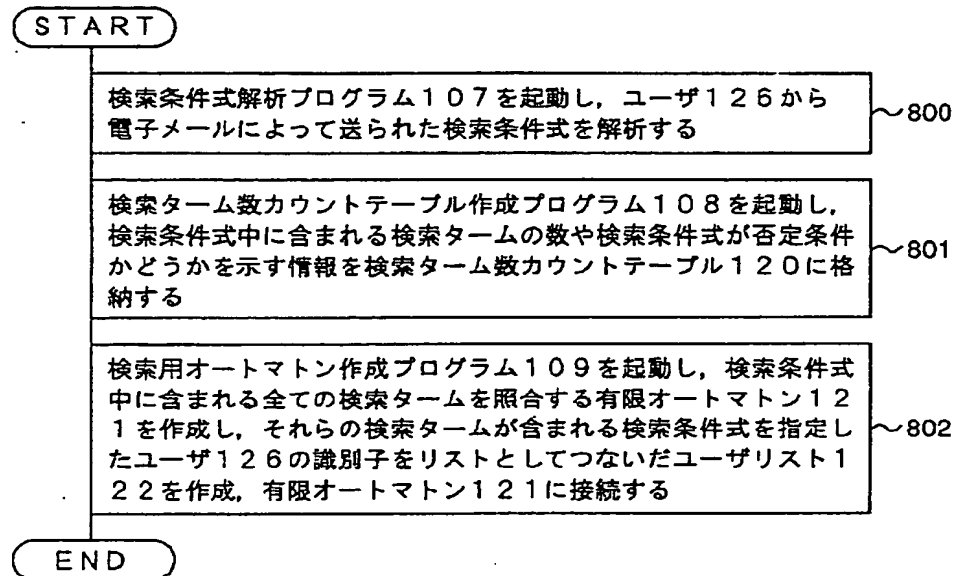


〈システム制御プログラム105の処理内容〉



【图8】

【图 8】



〈検索条件式登録制御プログラム106の処理内容〉

【圖 10】

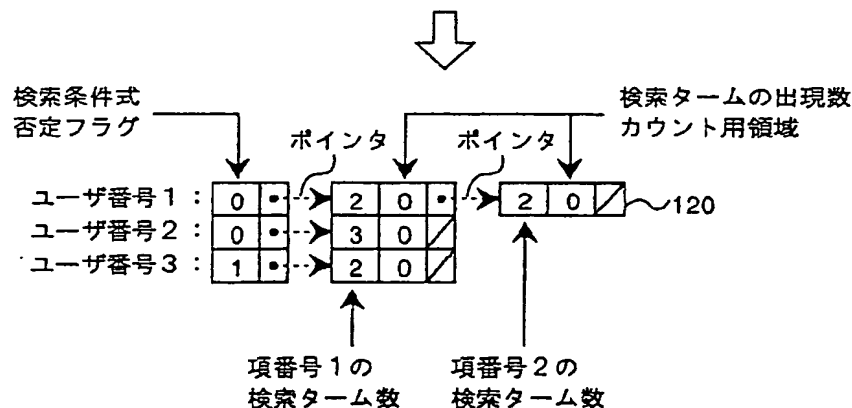
【图 10】

ユーザ番号1: 「(「文書」 and 「検索」) or (「文書」 and 「サーチ」)」

第1項	第2項
-----	-----

ユーザ番号2: 「“文字” and ¬ “認識” and ¬ “学習”」  
第1項

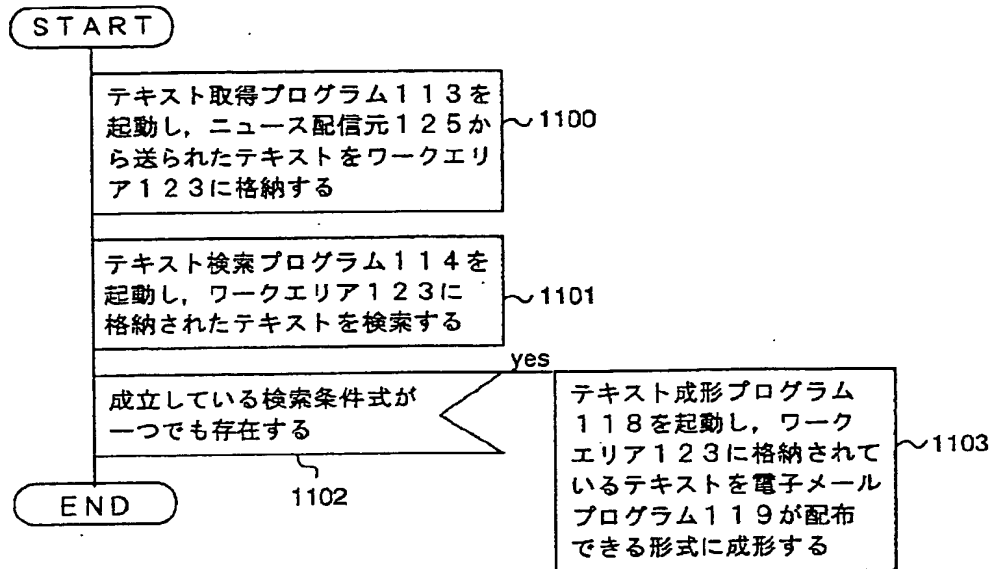
ユーザ番号3: 「「検索」 and 「学習」」  
第1項



〈検索ターム数カウントテーブル120の例〉

【図 11】

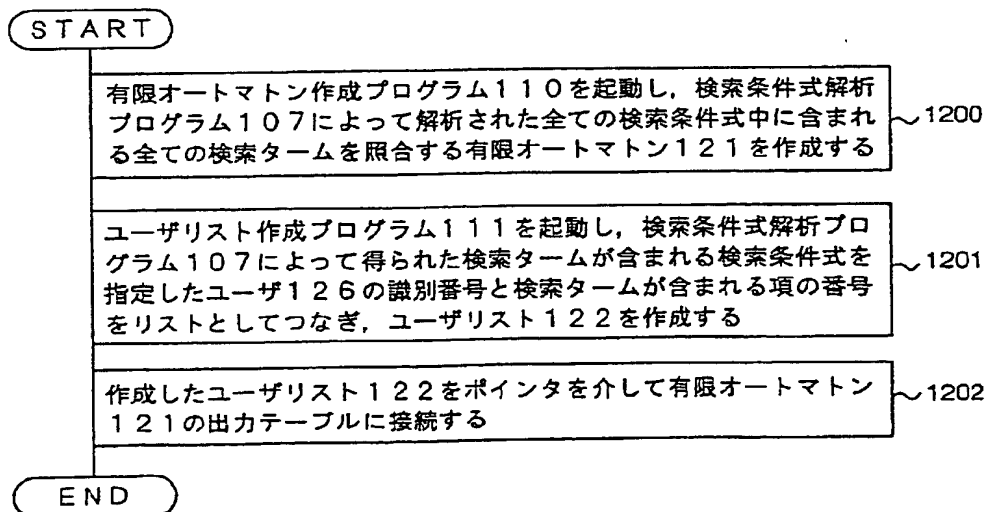
【図 11】



〈テキスト検索制御プログラム 112 の処理内容〉

【図 12】

【図 12】



〈検索用オートマトン作成プログラム 109 の処理内容〉

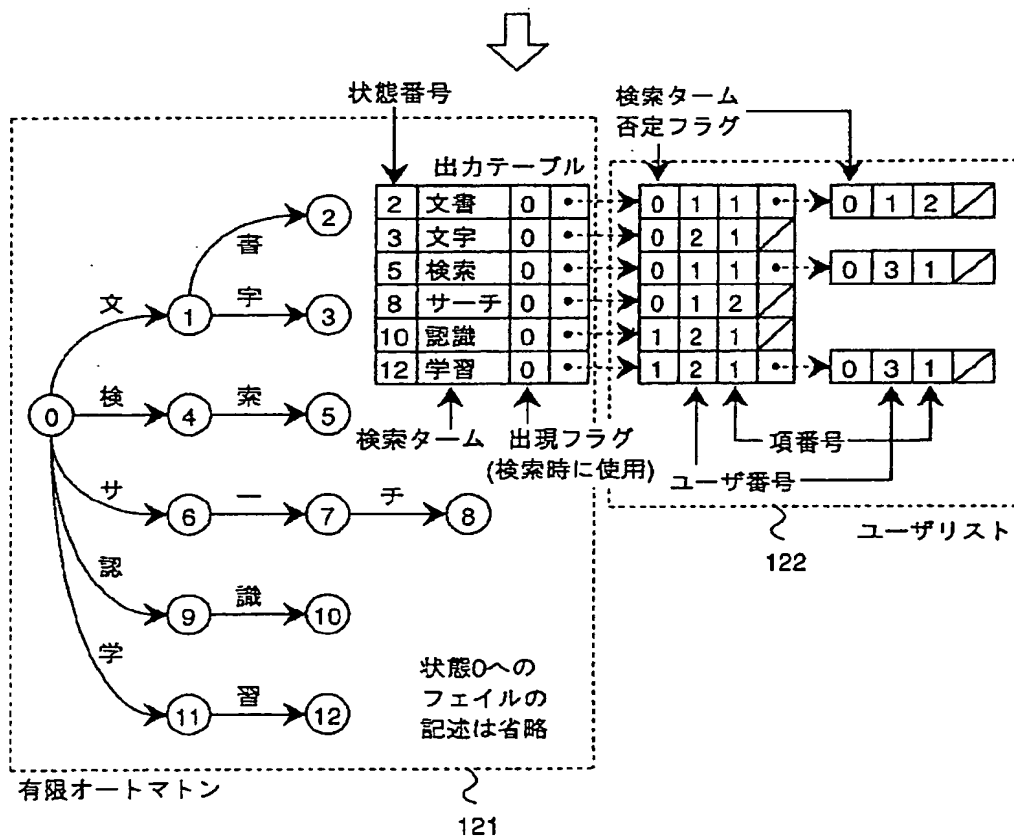
【図 13】

【図 13】

ユーザ番号 1 : 「 ( “文書” and “検索” ) or ( “文書” and “サーチ” ) 」  
 第 1 項 第 2 項

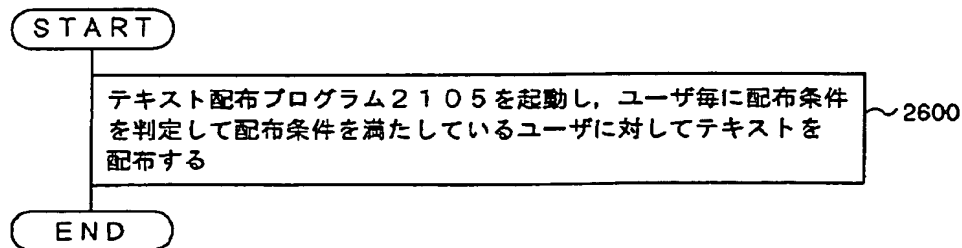
ユーザ番号 2 : 「 “文字” and ¬ “認識” and ¬ “学習” 」  
 第 1 項

ユーザ番号 3 : 「 ¬ ( “検索” and “学習” ) 」  
 第 1 項



【図 26】

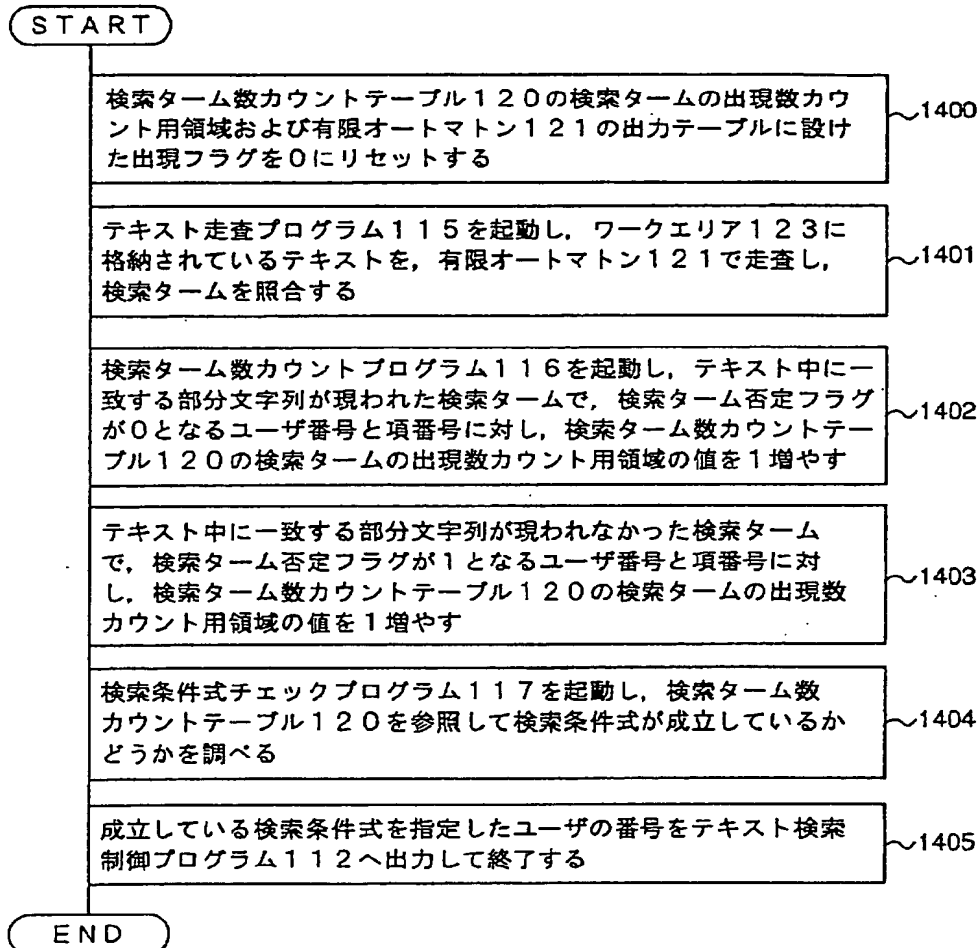
【図 26】



〈テキスト配布制御プログラム 2104 の処理内容〉

【図 14】

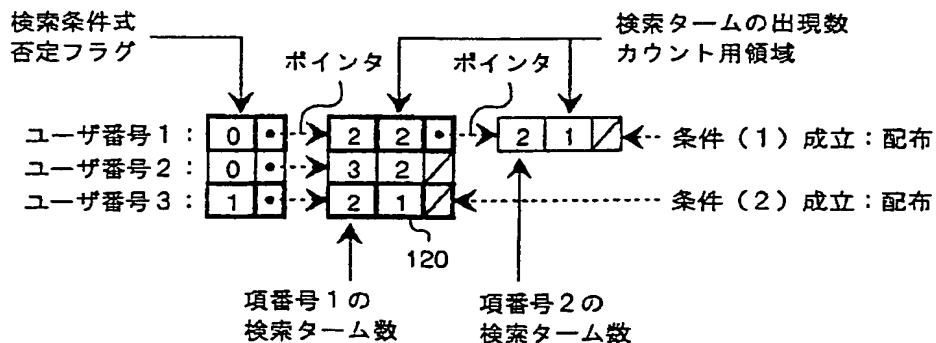
【図 14】



〈テキスト検索プログラム 114 の処理内容〉

【図 19】

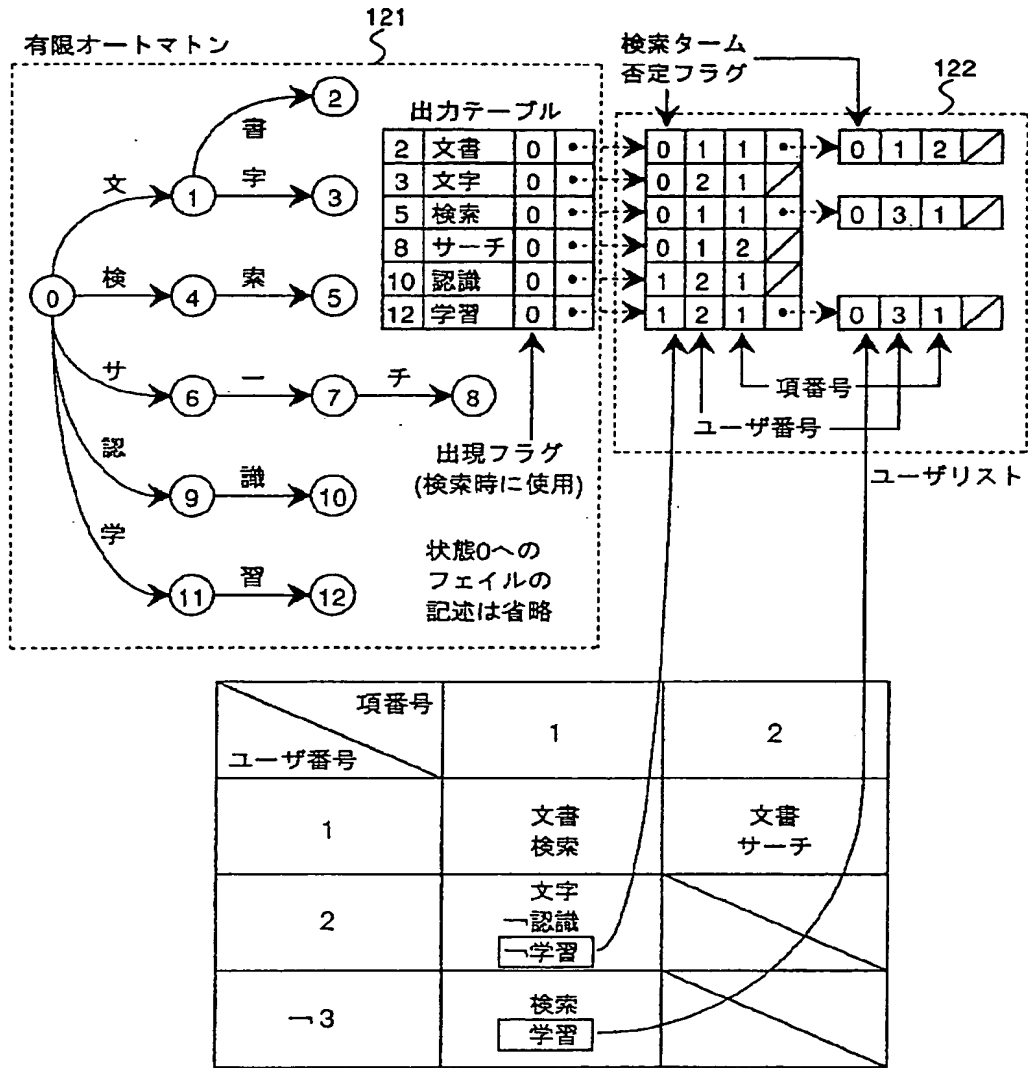
【図 19】



〈検索ターム数カウントテーブル 120 における検索条件式チェック処理の例〉

【図 15】

【図 15】



【図 36】

【図 36】

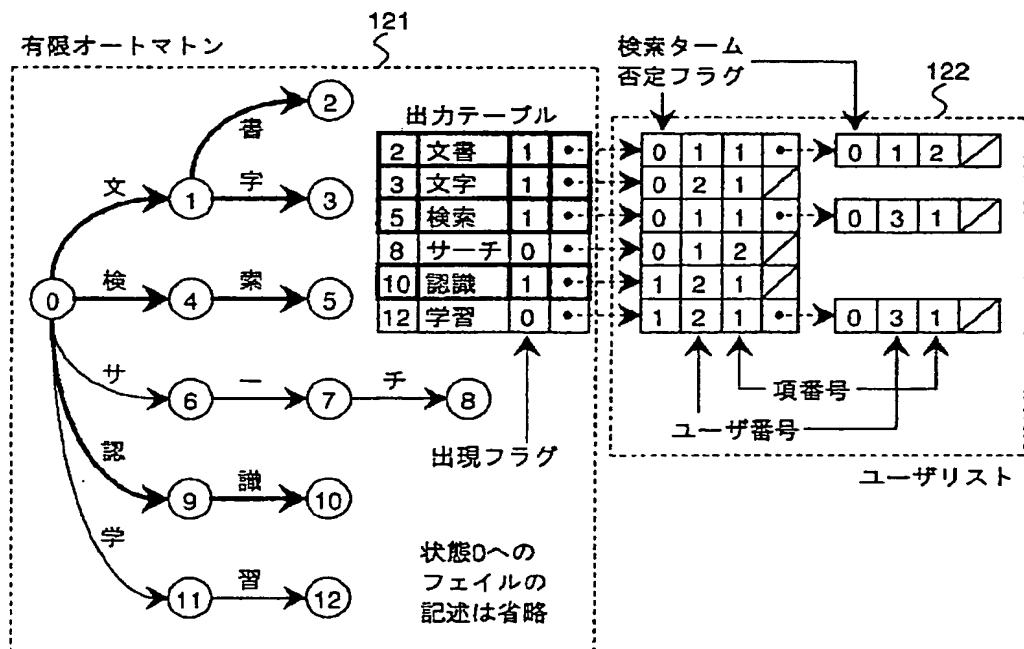
ユーザ番号 1	ユーザ番号 2	ユーザ番号 3	
文書 検索 サーチ	文字 認識 学習	検索 学習	3106

〈検索ターム管理テーブル 3106 の例〉

【図 16】

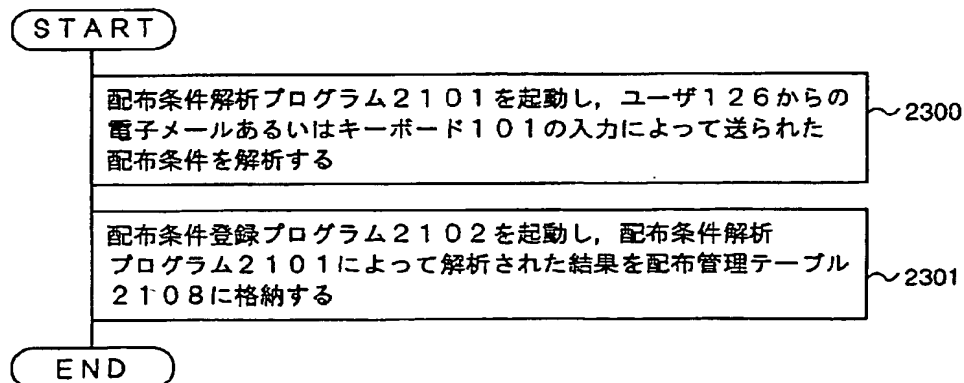
【図 16】

テキスト 検索した文書の書式を解析し、文字列部分を認識する



【図 23】

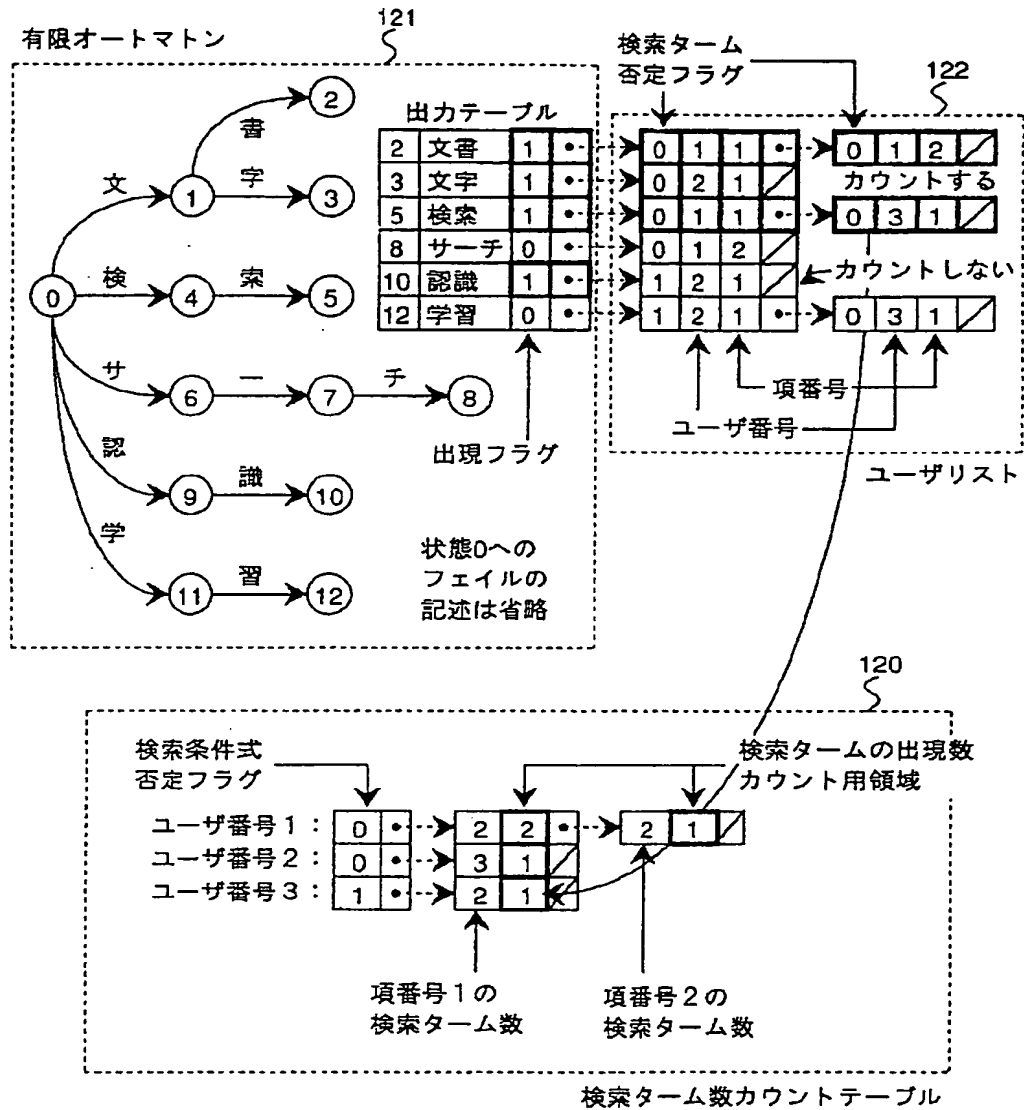
【図 23】



〈配布条件登録制御プログラム 2100 の処理内容〉

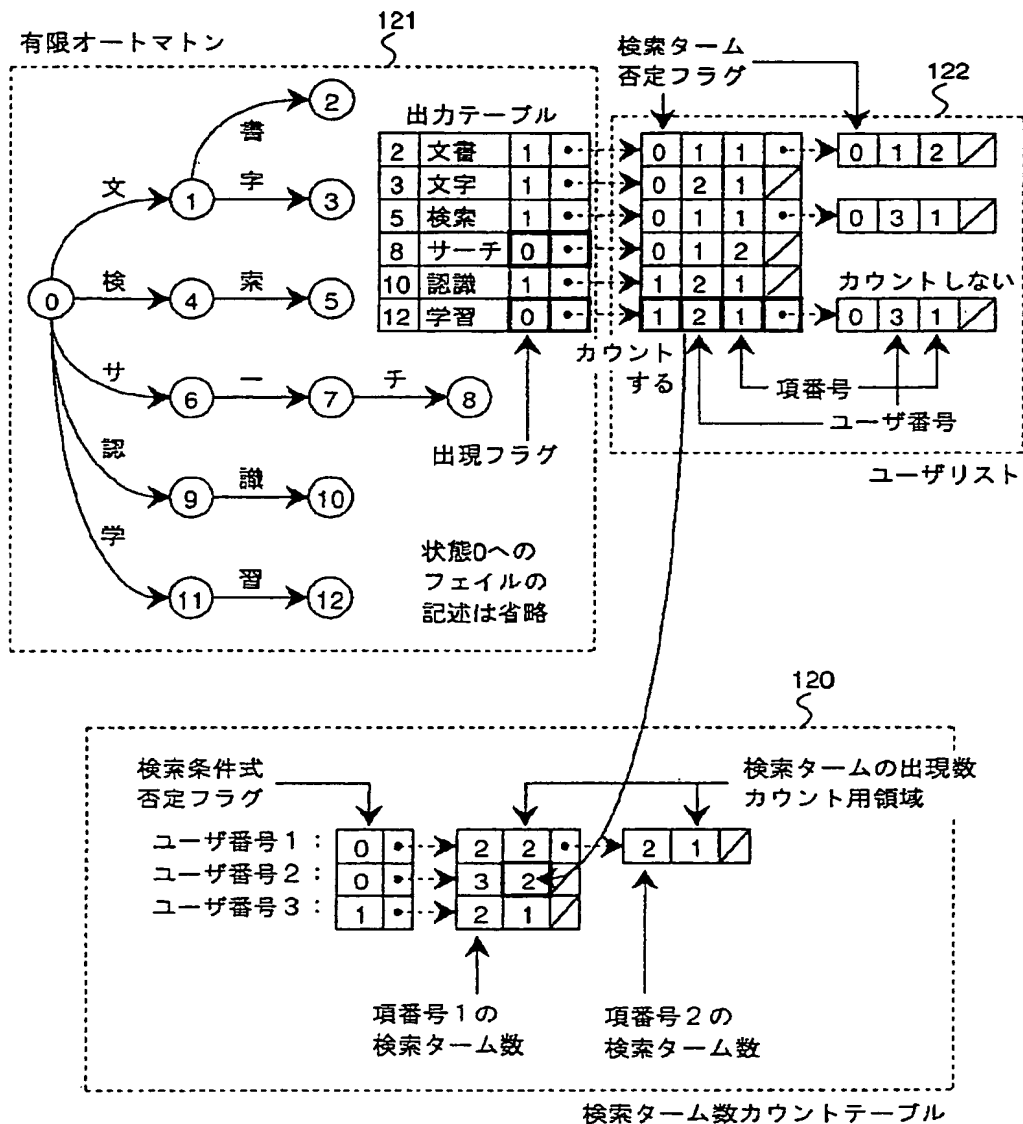
【図17】

【図17】



【図18】

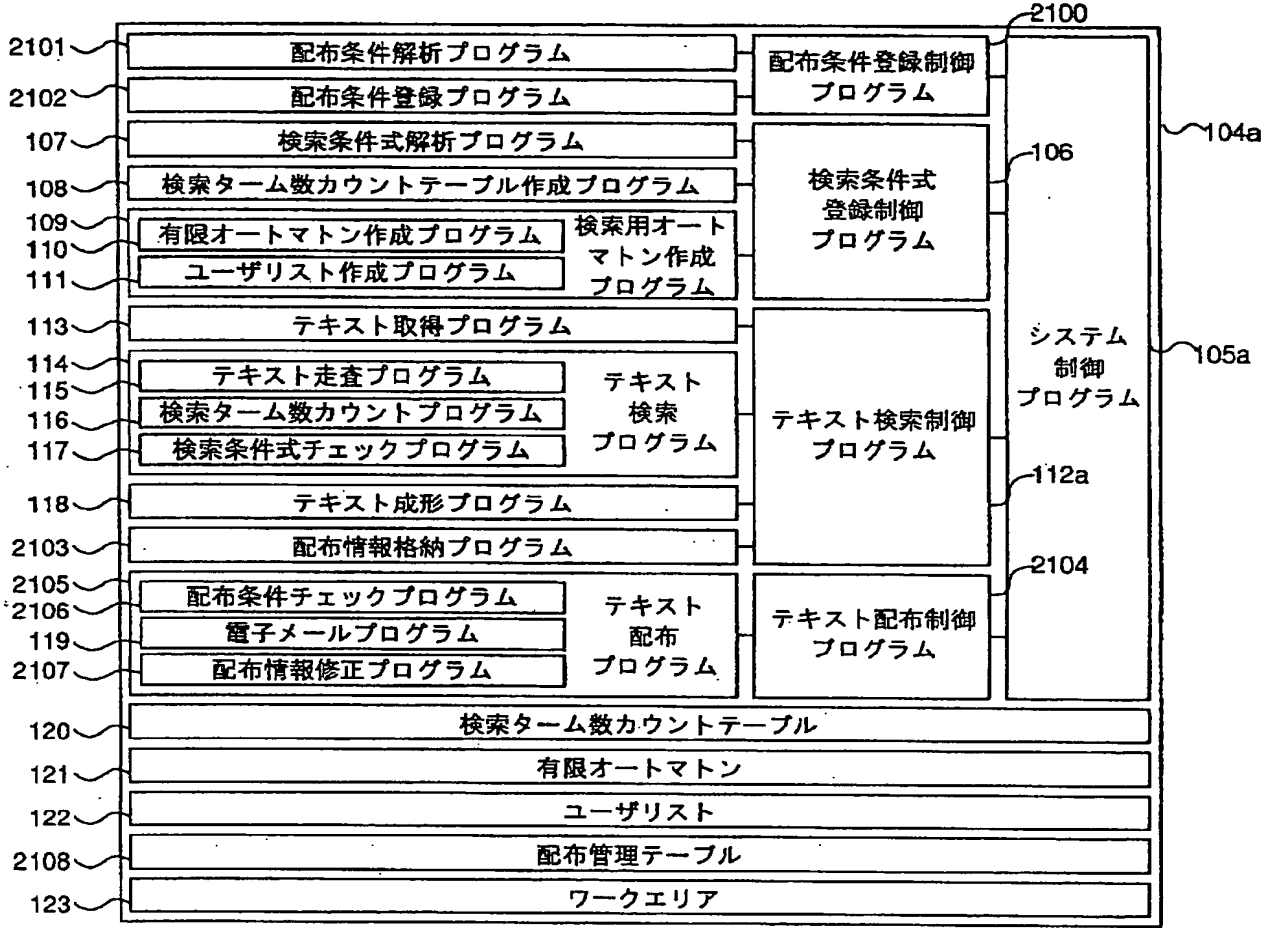
【図18】





【図 2 1】

【図 2 1】



【図 2 4】

【図 2 4】

配布条件の形式

配布条件チェック用領域

ユーザ番号 1 :	配布時間	18:00	NULL	NULL
ユーザ番号 2 :	配布件数	5	NULL	NULL
ユーザ番号 3 :	遅延時間	01:30	NULL	NULL

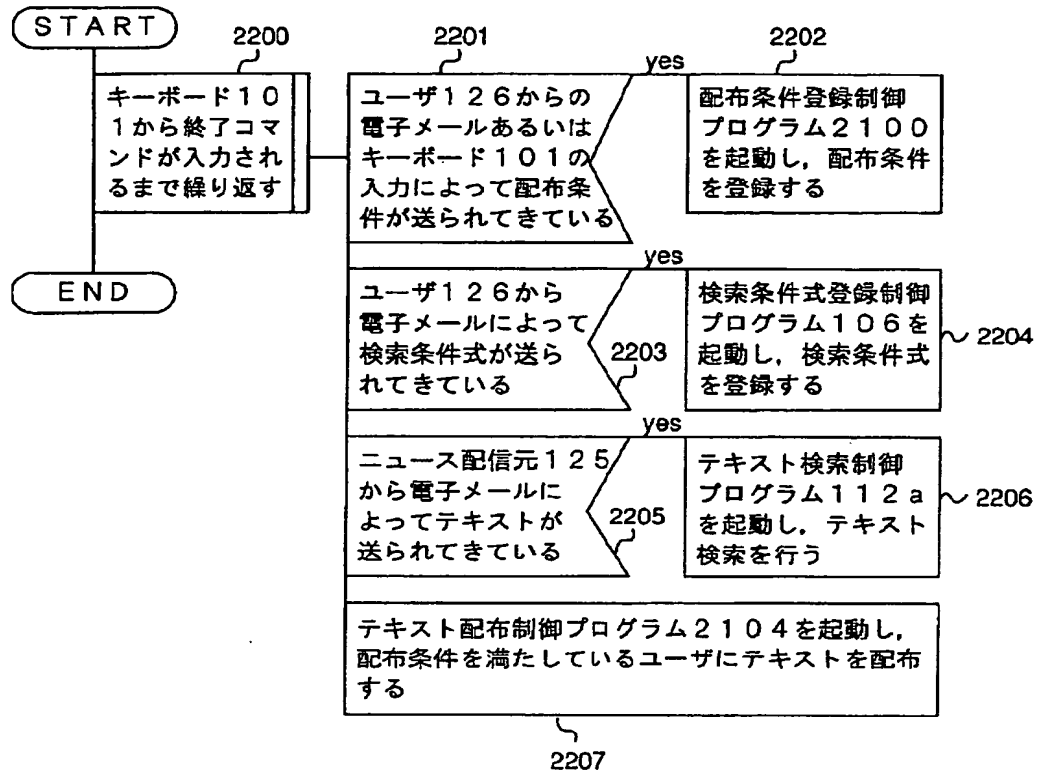
配布条件の設定値

配布テキスト番号格納用領域

〈配布管理テーブル 2108 の例〉

【図 22】

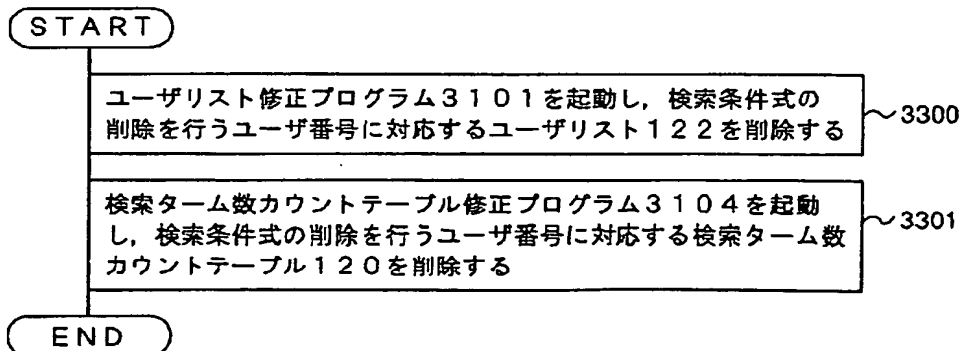
【図 22】



〈システム制御プログラム105aの処理内容〉

【図 33】

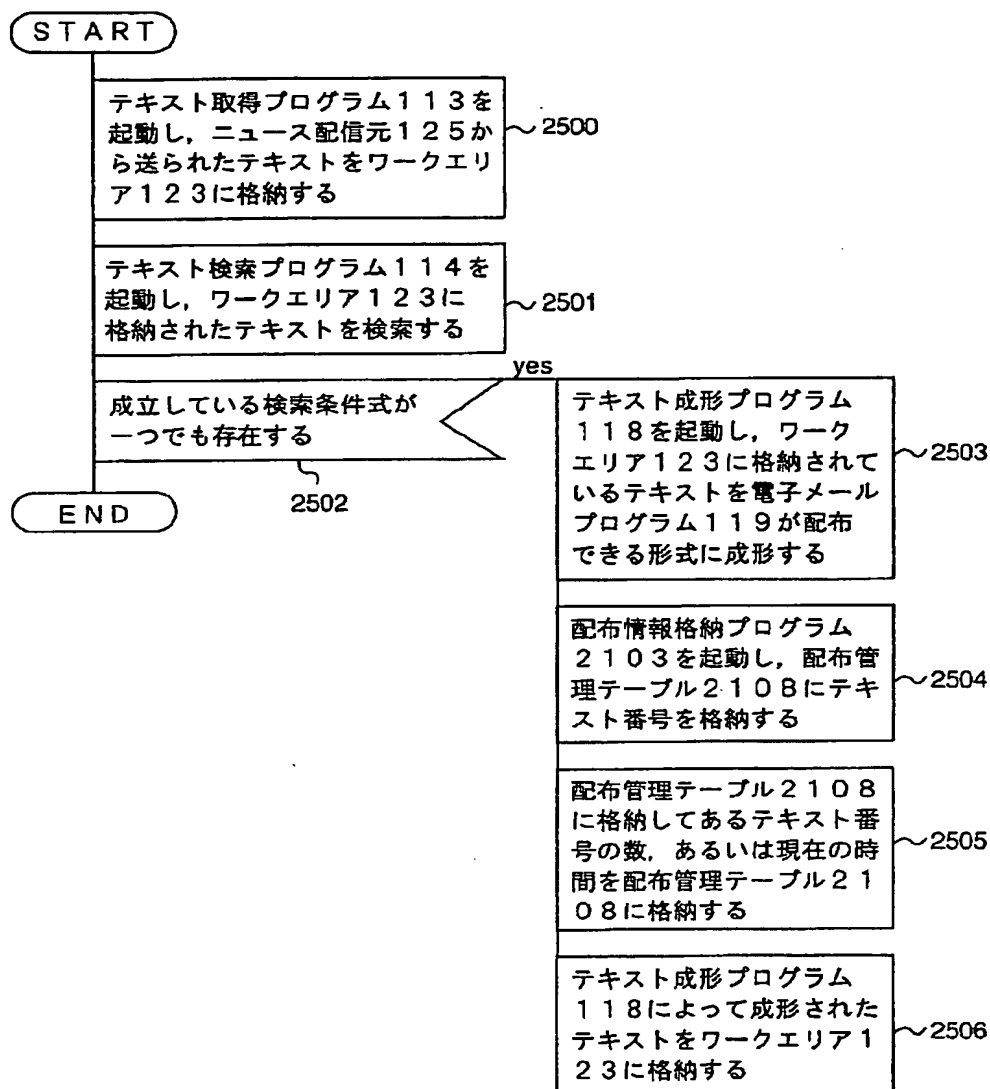
【図 33】



〈検索条件式削除制御プログラム3100の処理内容〉

【図 25】

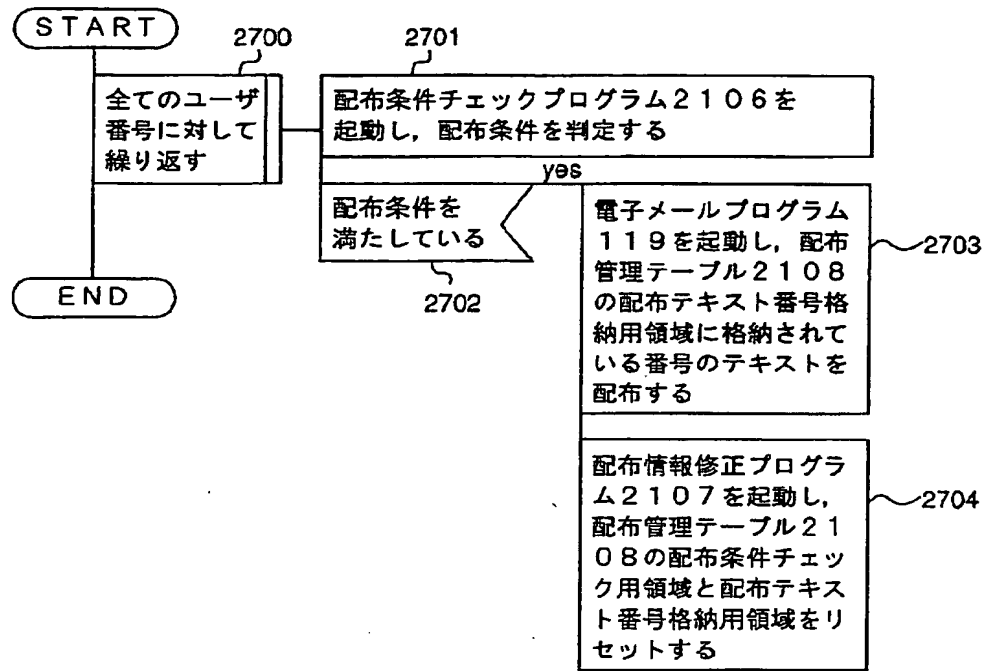
【図 25】



〈テキスト検索制御プログラム112aの処理内容〉

【図27】

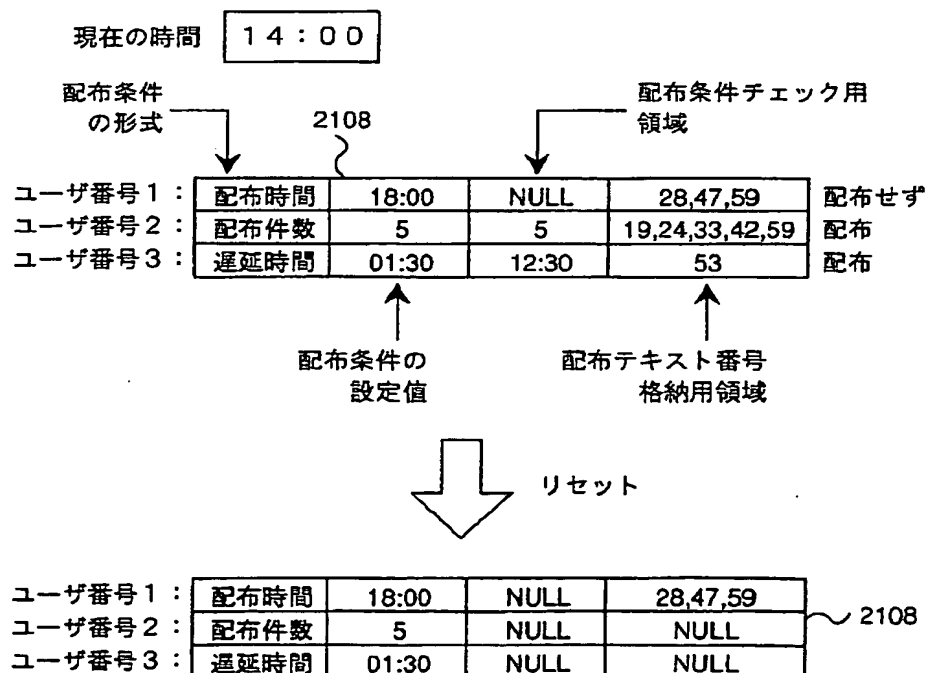
【図27】



〈テキスト配布プログラム2105の処理内容〉

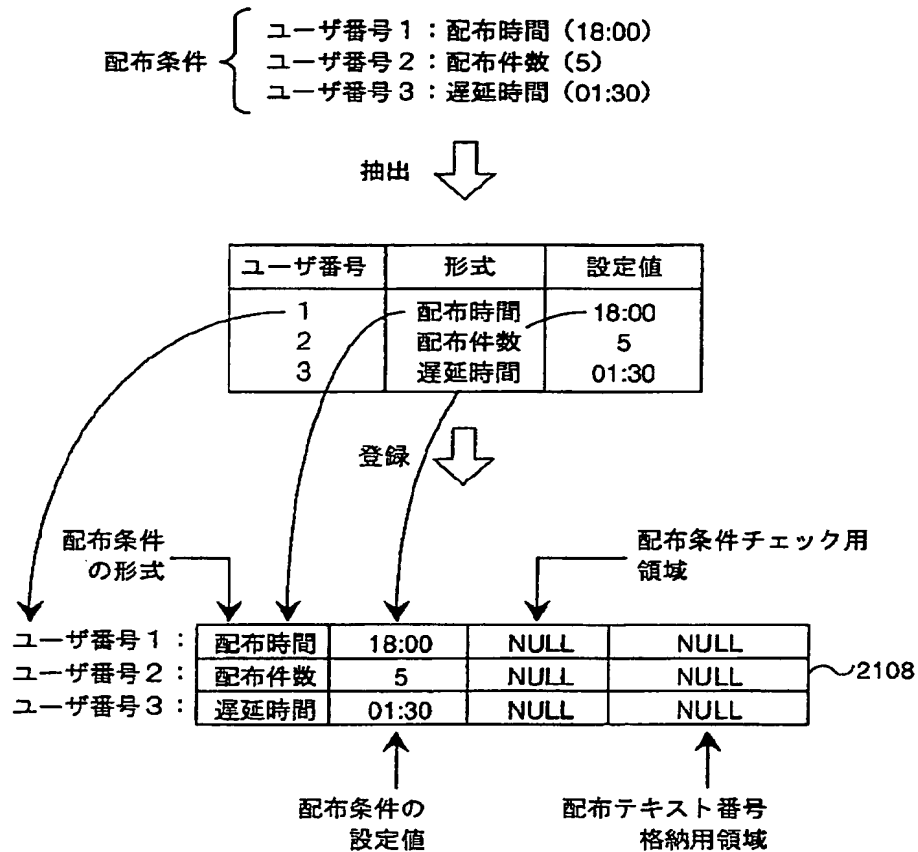
【図30】

【図30】



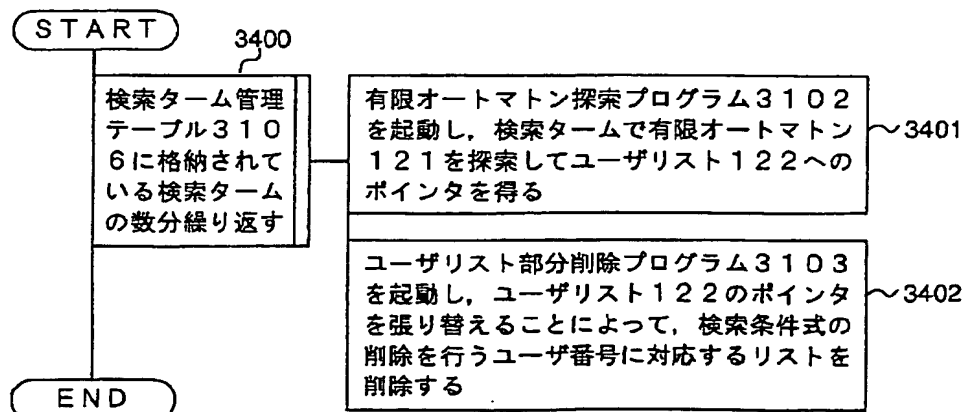
【図 28】

【図 28】



【図 34】

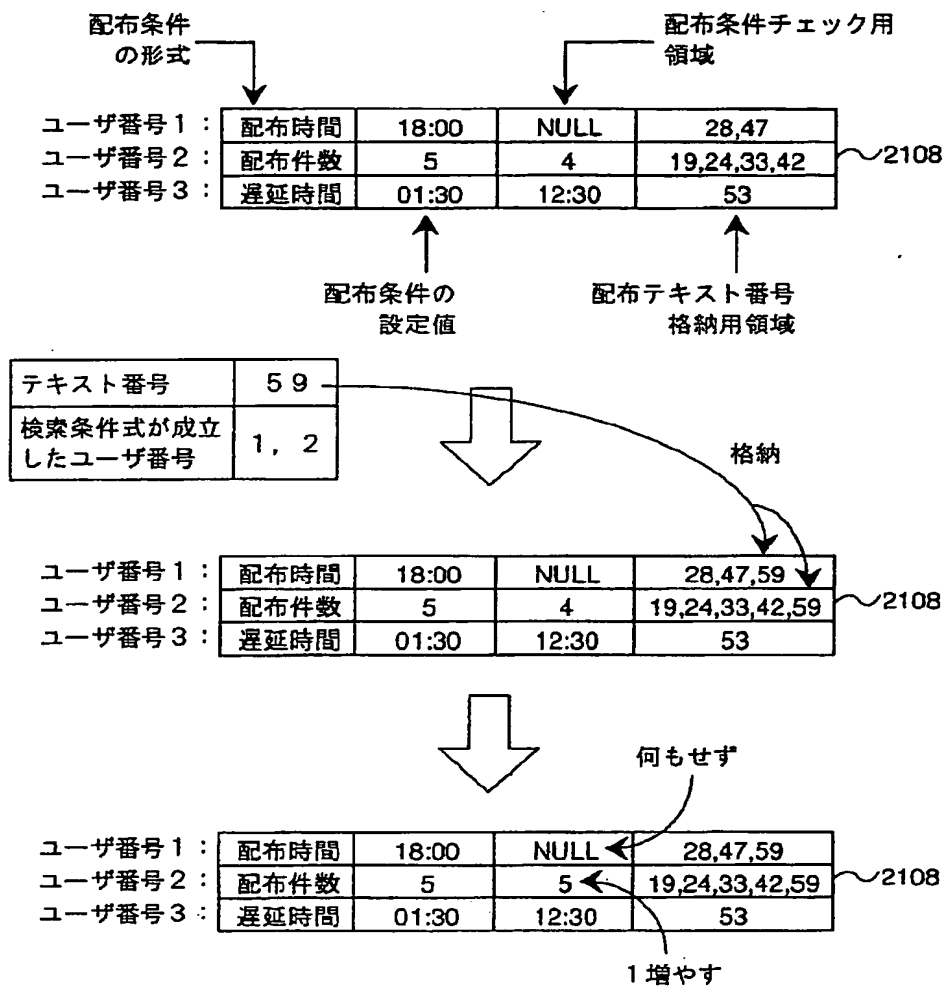
【図 34】



〈ユーザリスト修正プログラム 3101 の処理内容〉

【図29】

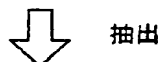
【図29】



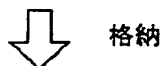
【図39】

【図39】

ユーザ番号2: 「“構造” and “認識”」



「構造」, 「認識」

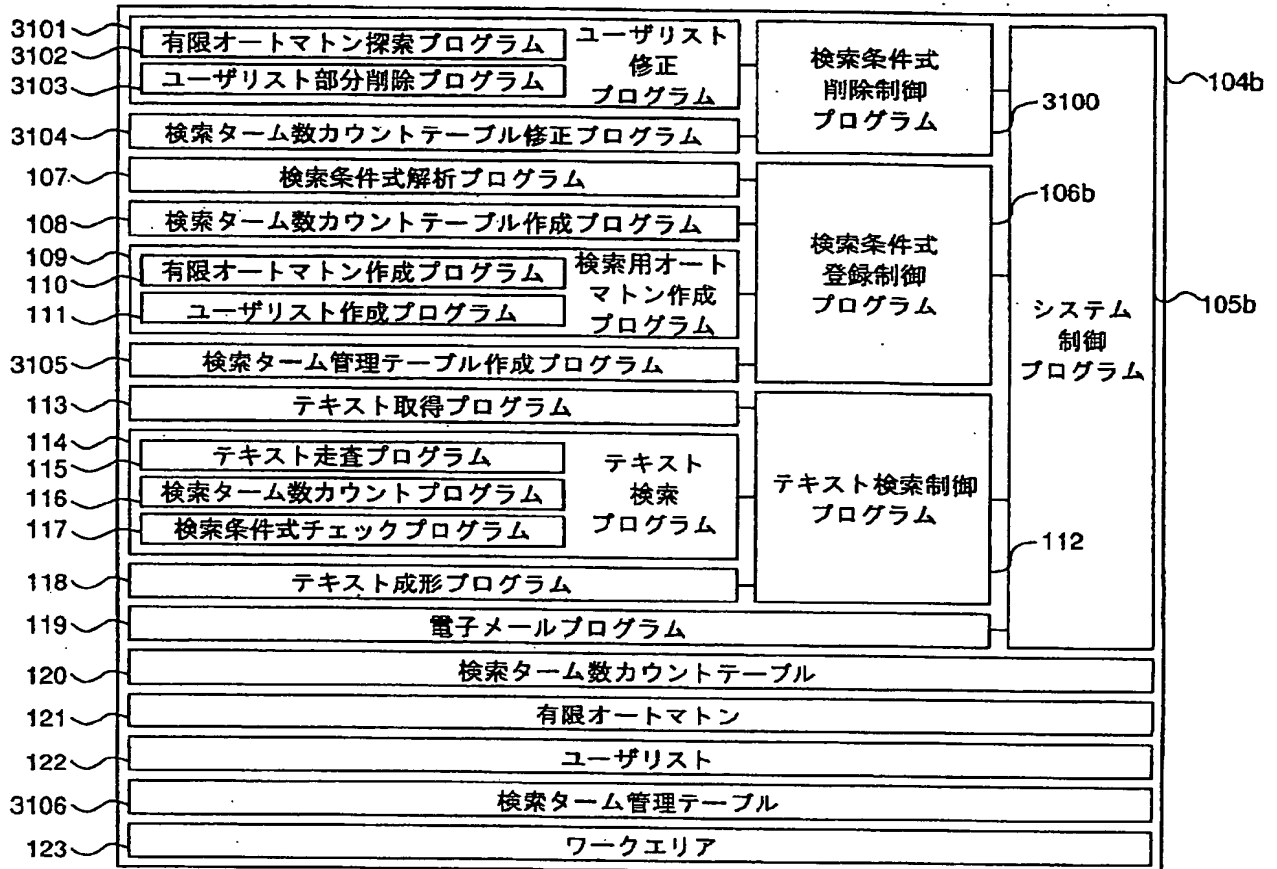


ユーザ番号1	ユーザ番号2	ユーザ番号3
文書 検索 サーチ	構造 認識	検索 学習

3106

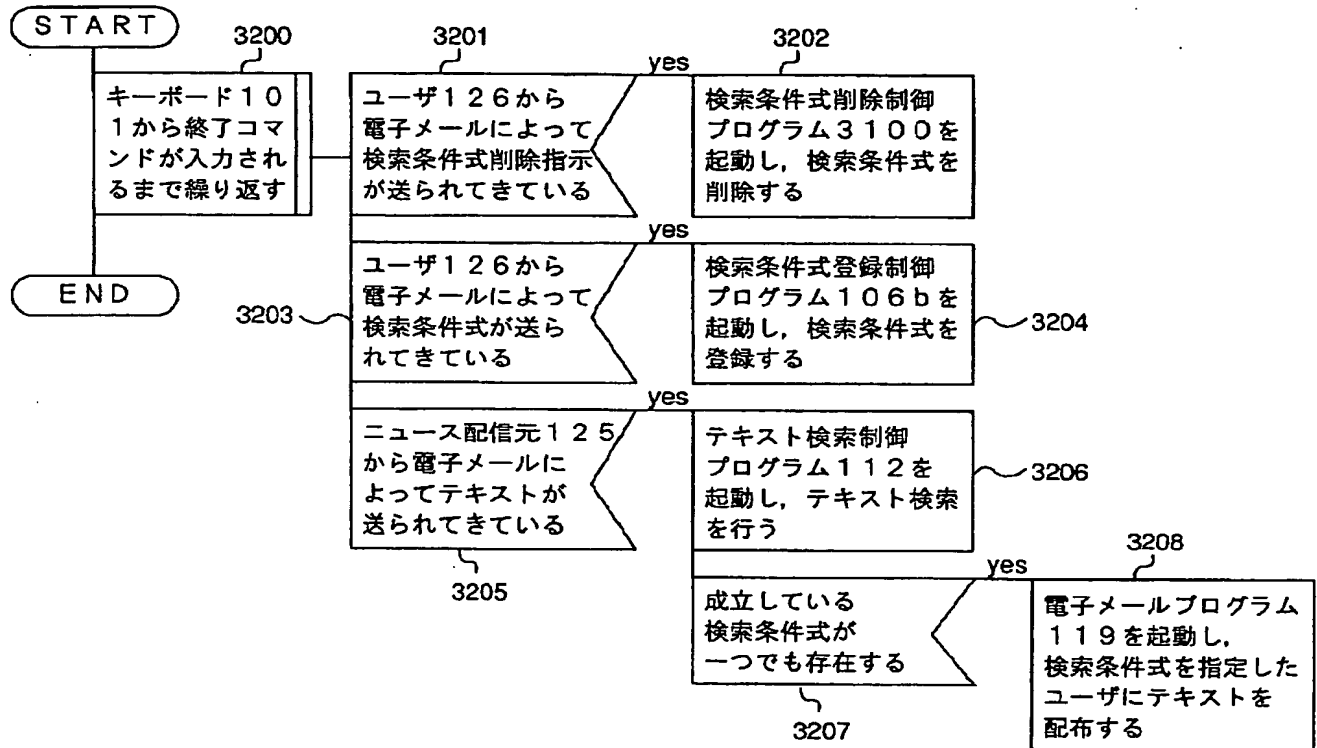
【図 31】

【図 31】



【図 32】

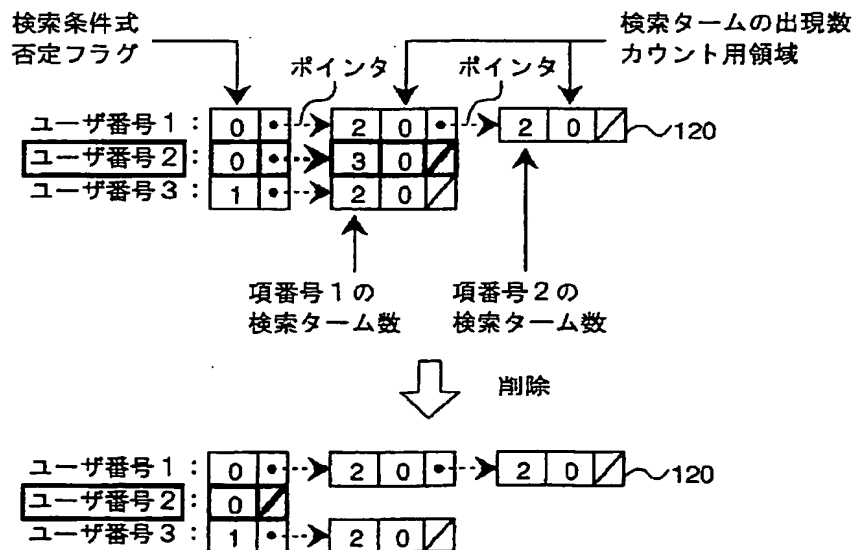
【図 32】



(システム制御プログラム105bの処理内容)

【図 38】

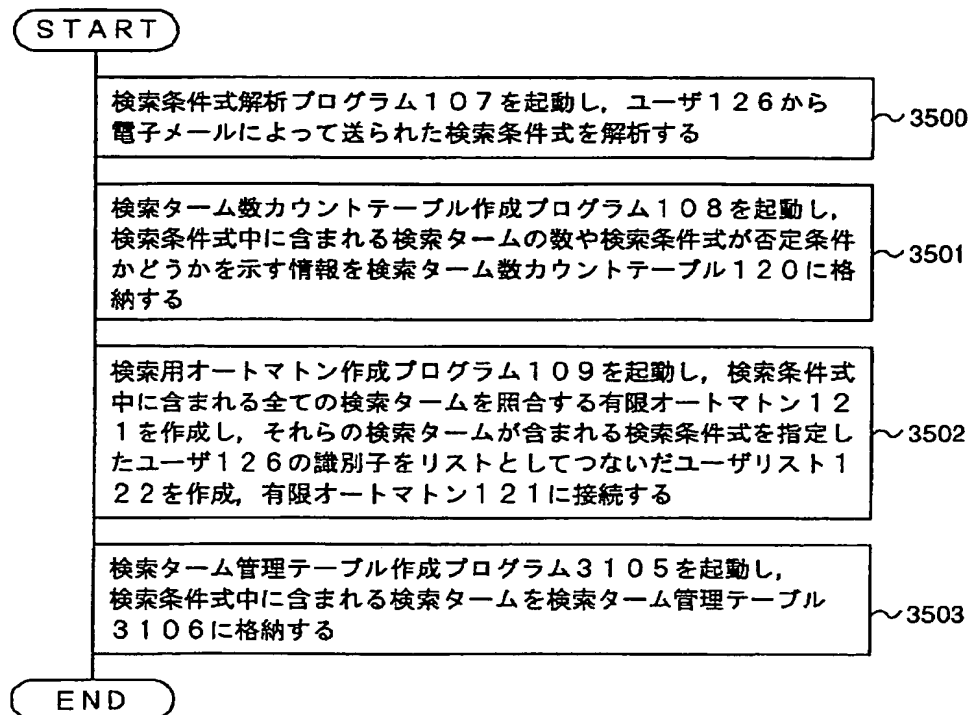
【図 38】





【図 35】

【図 35】



〈検索条件式登録制御プログラム106bの処理内容〉

【図37】

【図37】

